

Intro to Data Visualisation

Dr Geraint Palmer-Liyu

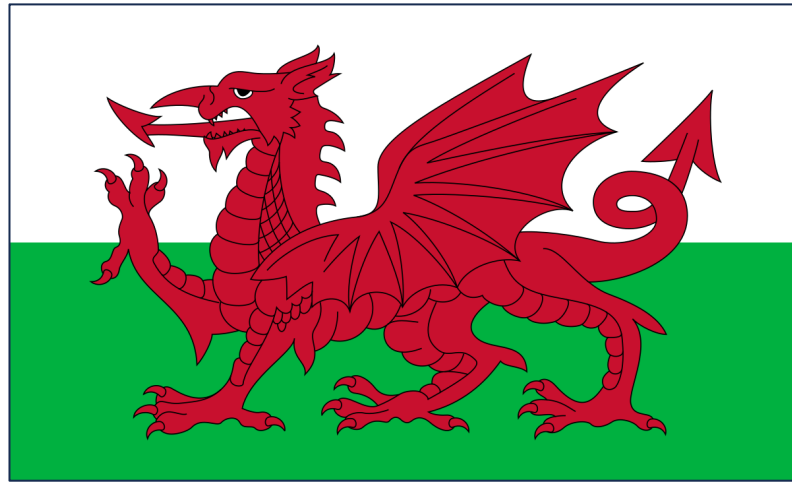
School of Mathematics - Cardiff University

FEB UBB – Cardiff University: International Visiting Lecture Series

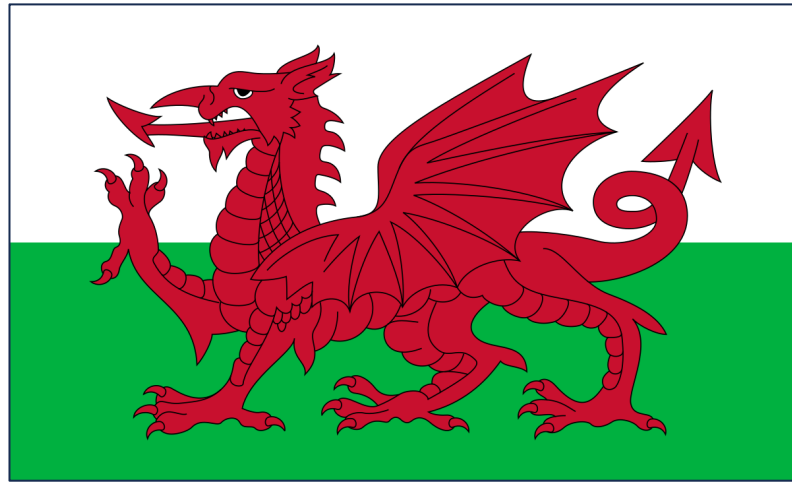
Who are we?



Who are we?



Who are we?



CARDIFF
UNIVERSITY

PRIFYSGOL
CAERDYDD



*Ar lan y môr mae rhosys cochion,
Ar lan y môr mae lilis gwynion,
Ar lan y môr mae 'nghariad inne
Yn cysgu'r nos a chodi'r bore.*



Structure

- 1.The what and the why
- 2.Comparing 2 or more variables
- 3.Exploring 1 variable
- 4.Common Pitfalls

1) The What and the Why

1. Why Visualise? Anscombe's Quartet

1. Why Visualise? Anscombe's Quartet

Height	Weight
10	8.04
8	6.95
13	7.58
9	8.81
11	8.33
14	9.96
6	7.24
4	4.26
12	10.84
7	4.82
5	5.68

Height	Weight
10	9.14
8	8.14
13	8.74
9	8.77
11	9.26
14	8.10
6	6.13
4	3.10
12	9.13
7	7.26
5	4.74

Height	Weight
10	7.46
8	6.77
13	12.74
9	7.11
11	7.81
14	8.84
6	6.08
4	5.39
12	8.15
7	6.42
5	5.73

Height	Weight
8	6.58
8	7.76
8	7.71
8	8.84
8	8.47
8	7.04
8	5.25
19	12.50
8	5.56
8	7.91
8	6.89

1. Why Visualise? Anscome's Quartet

Height	Weight
10	8.04
8	6.95
13	7.58
9	8.81
11	8.33
14	9.96
6	7.24
4	4.26
12	10.84
7	4.82
5	5.68

Height	Weight
10	9.14
8	8.14
13	8.74
9	8.77
11	9.26
14	8.10
6	6.13
4	3.10
12	9.13
7	7.26
5	4.74

Height	Weight
10	7.46
8	6.77
13	12.74
9	7.11
11	7.81
14	8.84
6	6.08
4	5.39
12	8.15
7	6.42
5	5.73

Height	Weight
8	6.58
8	7.76
8	7.71
8	8.84
8	8.47
8	7.04
8	5.25
19	12.50
8	5.56
8	7.91
8	6.89

Mean height = 9

Mean weight = 7.5

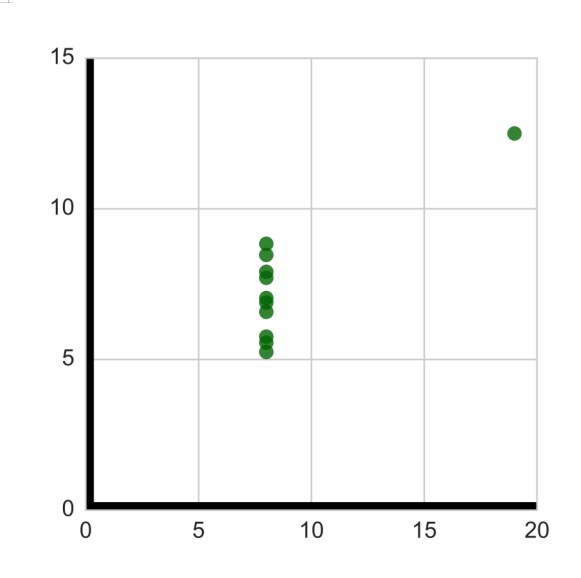
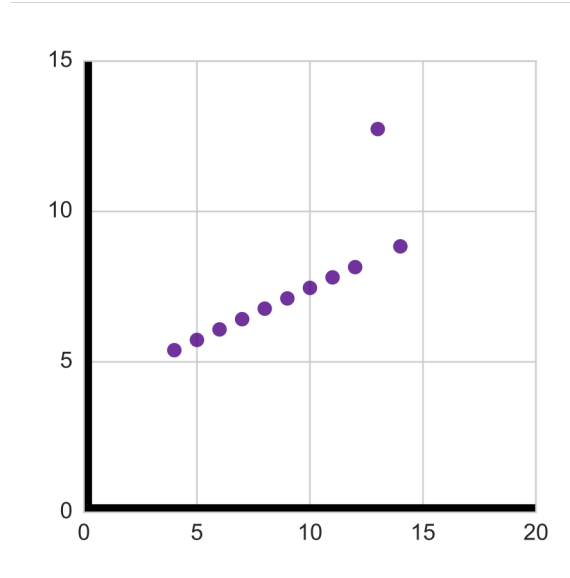
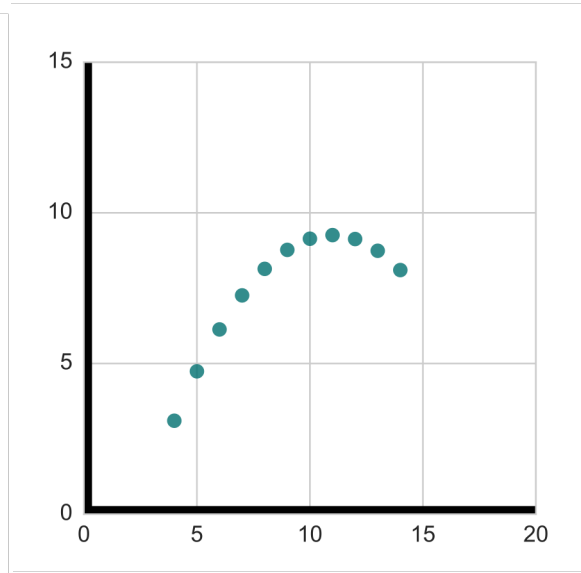
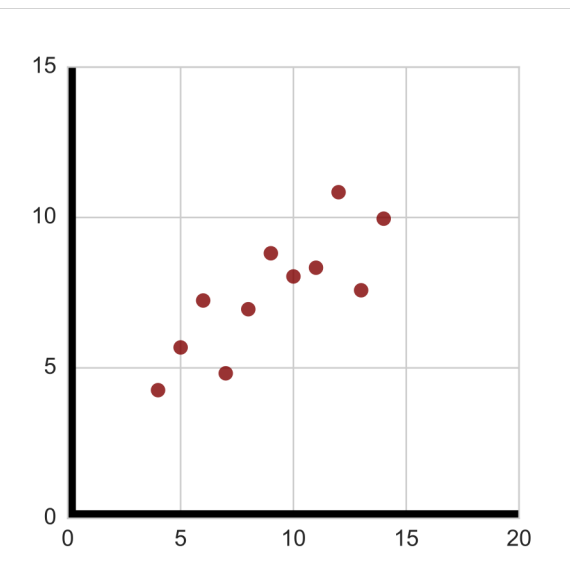
Var height = 11

Var weight = 4.12

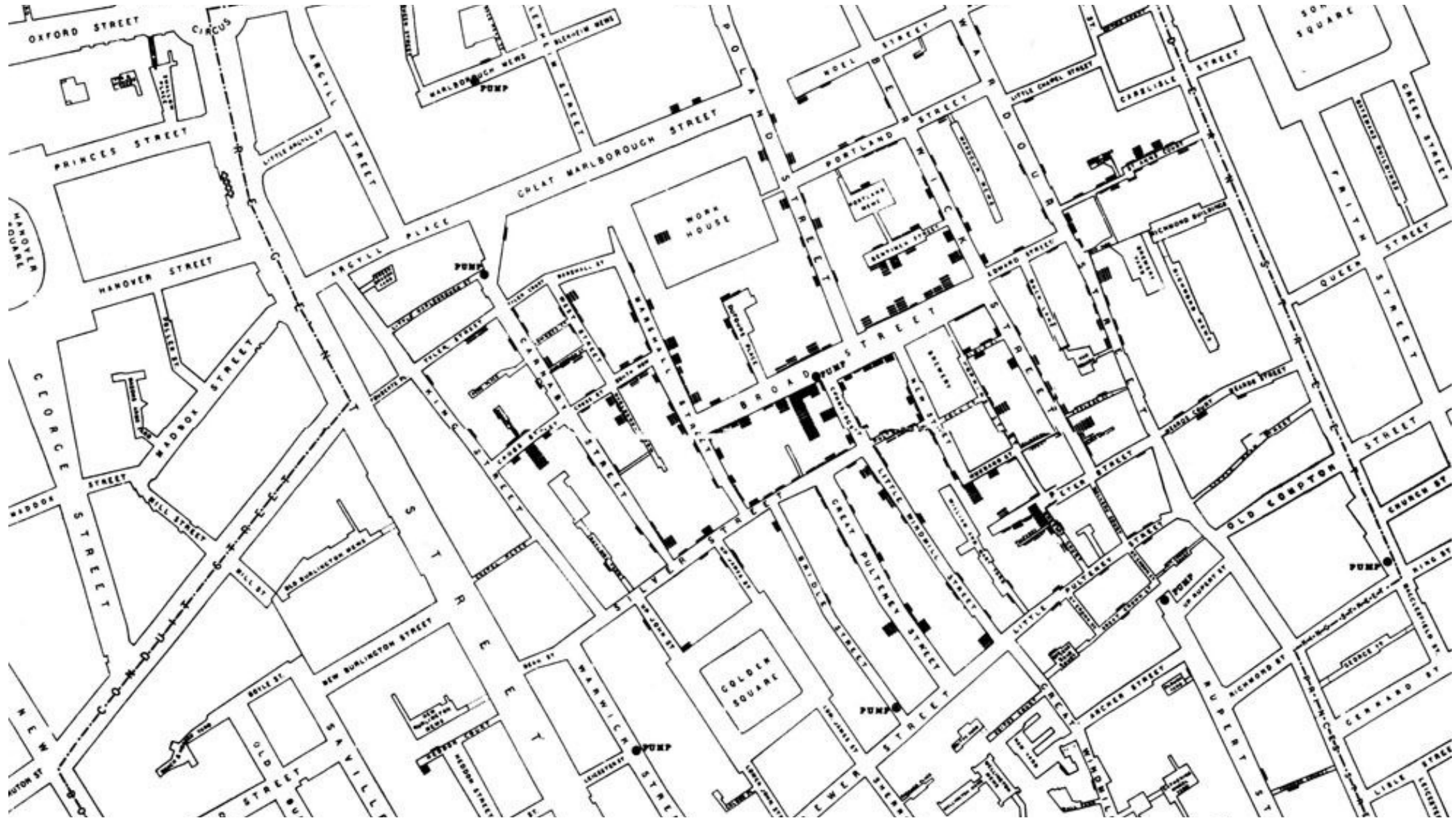
Correlation = 0.816

Line of best fit: $W = 0.5H + 3$

1. Why Visualise? Anscombe's Quartet



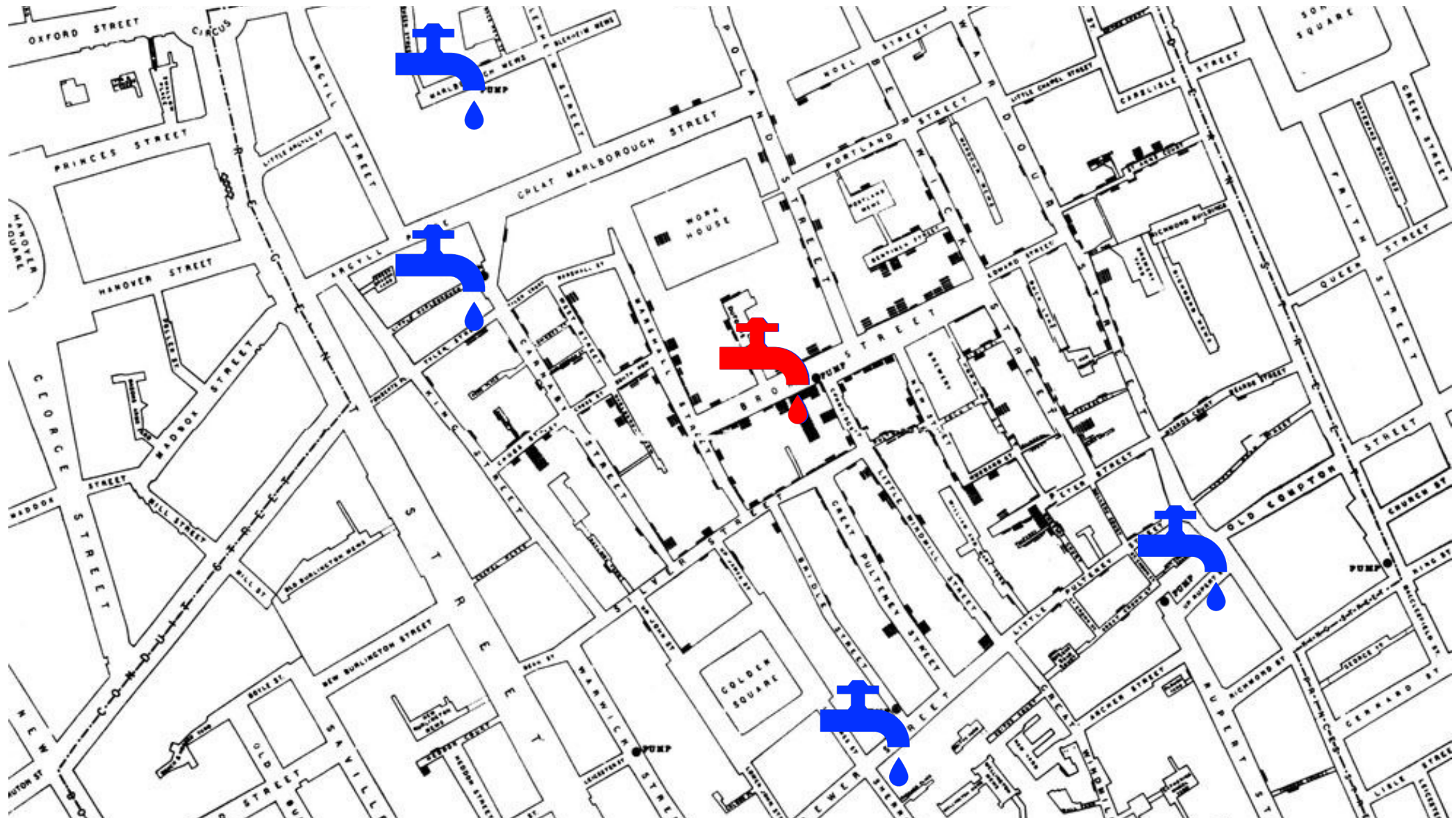
1. History – John Snow Cholera Map (1854)



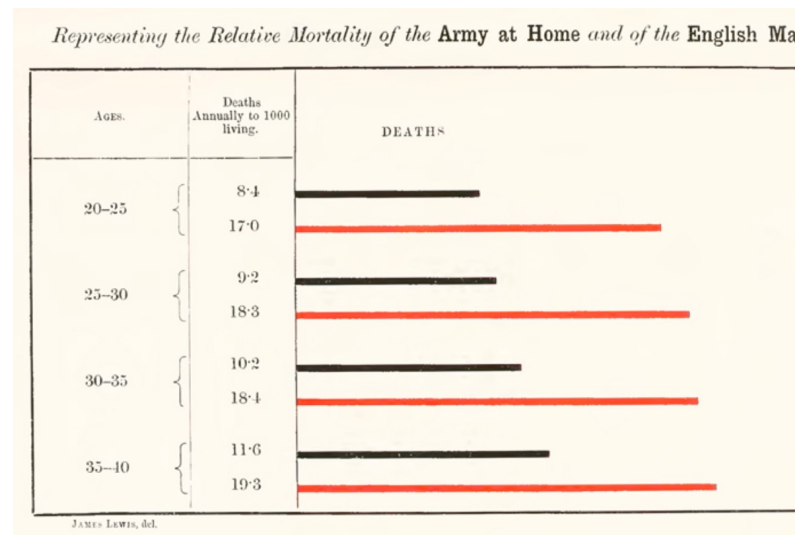
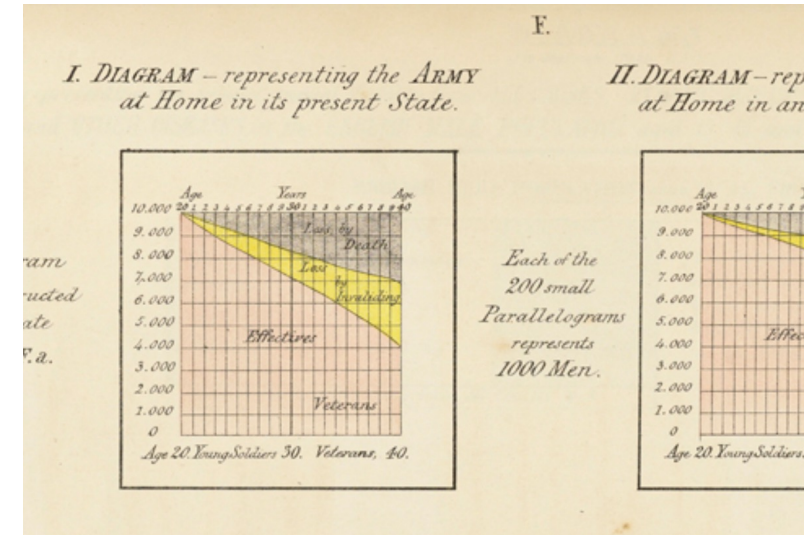
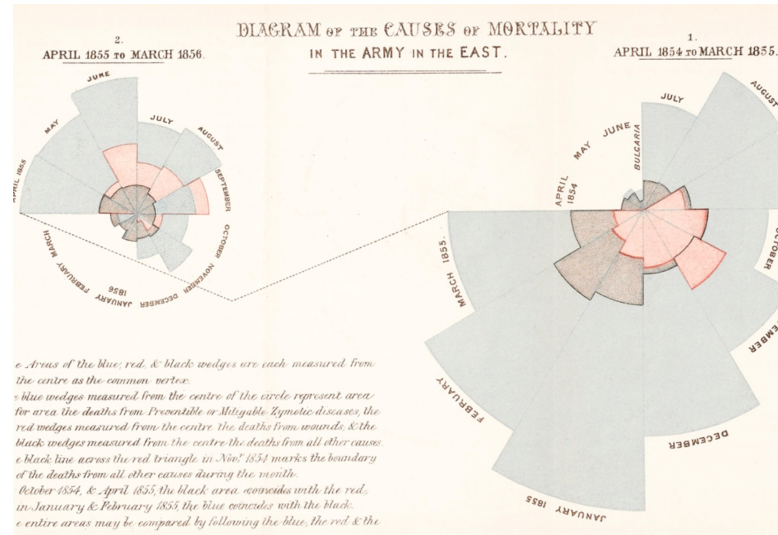
1. History – John Snow Cholera Map (1854)



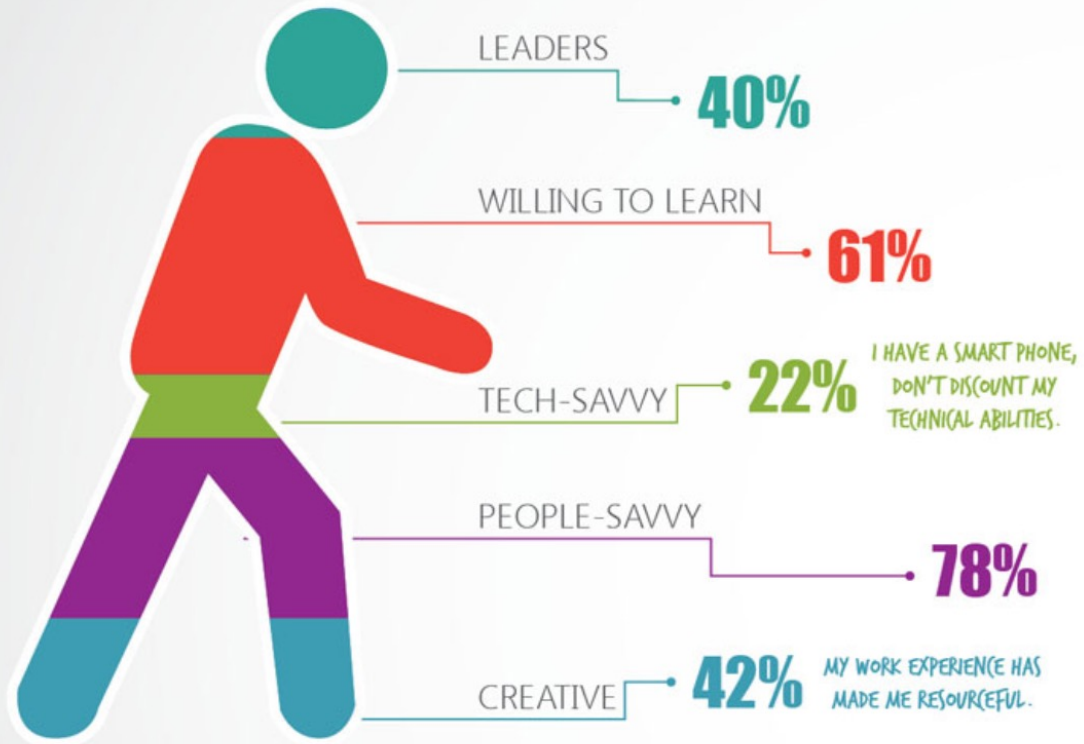
1. History – John Snow Cholera Map (1854)



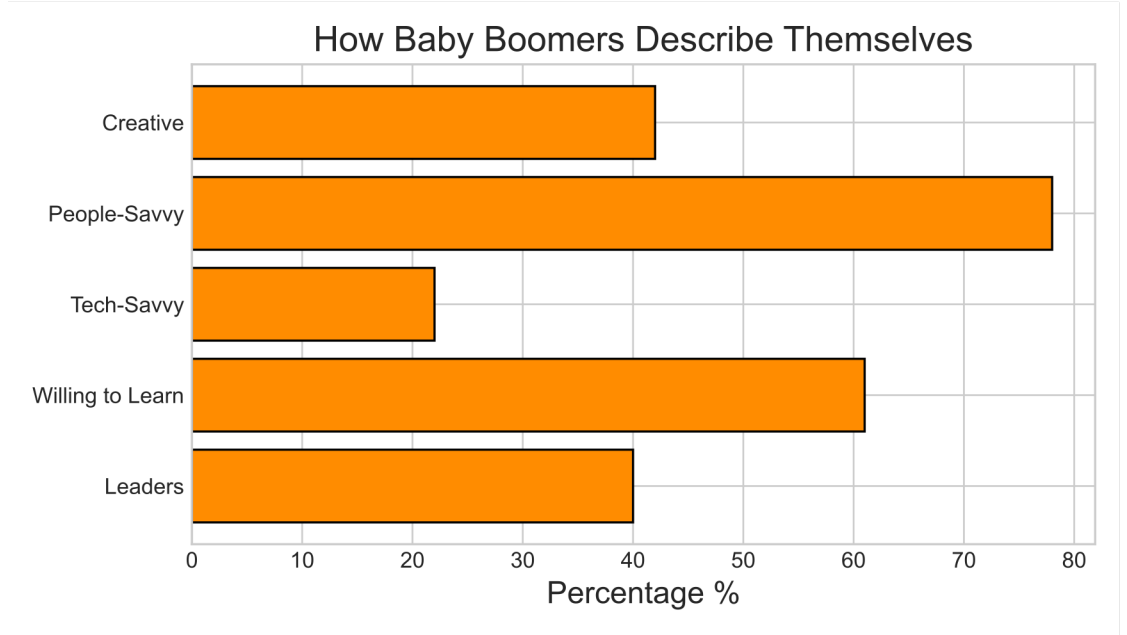
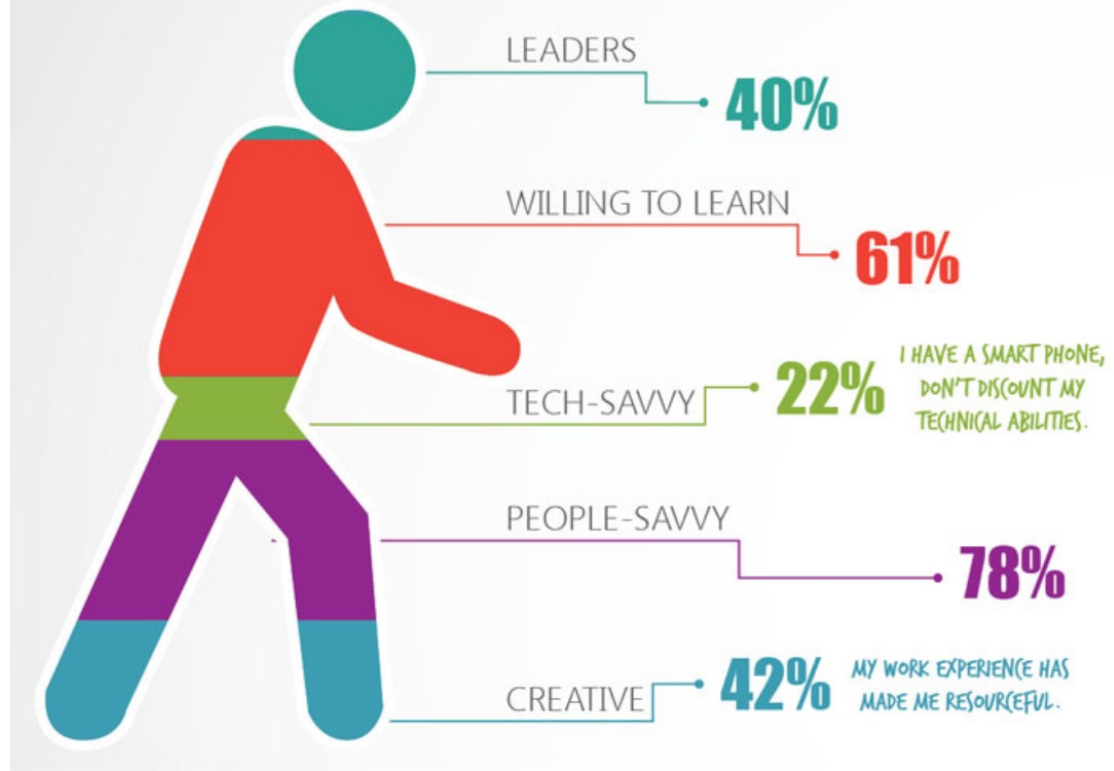
1. History – Florence Nightingale (1858)



HOW BABY BOOMERS DESCRIBE THEMSELVES



HOW BABY BOOMERS DESCRIBE THEMSELVES



1. Data Visualisation is not Art

Art

Data Visualisation

1. Data Visualisation is not Art

Art

Data Visualisation

Purpose

Evokes emotions and persuades

Communicates data and illuminates patterns

1. Data Visualisation is not Art

	Art	Data Visualisation
Purpose	Evokes emotions and persuades	Communicates data and illuminates patterns
Point of View	Artist's point of view	No point of view

1. Data Visualisation is not Art

	Art	Data Visualisation
Purpose	Evokes emotions and persuades	Communicates data and illuminates patterns
Point of View	Artist's point of view	No point of view
Prescriptiveness	Open to interpretation	Unambiguous

1. Data Visualisation is not Art

	Art	Data Visualisation
Purpose	Evokes emotions and persuades	Communicates data and illuminates patterns
Point of View	Artist's point of view	No point of view
Prescriptiveness	Open to interpretation	Unambiguous
Approach	Free	Prescribed by the data

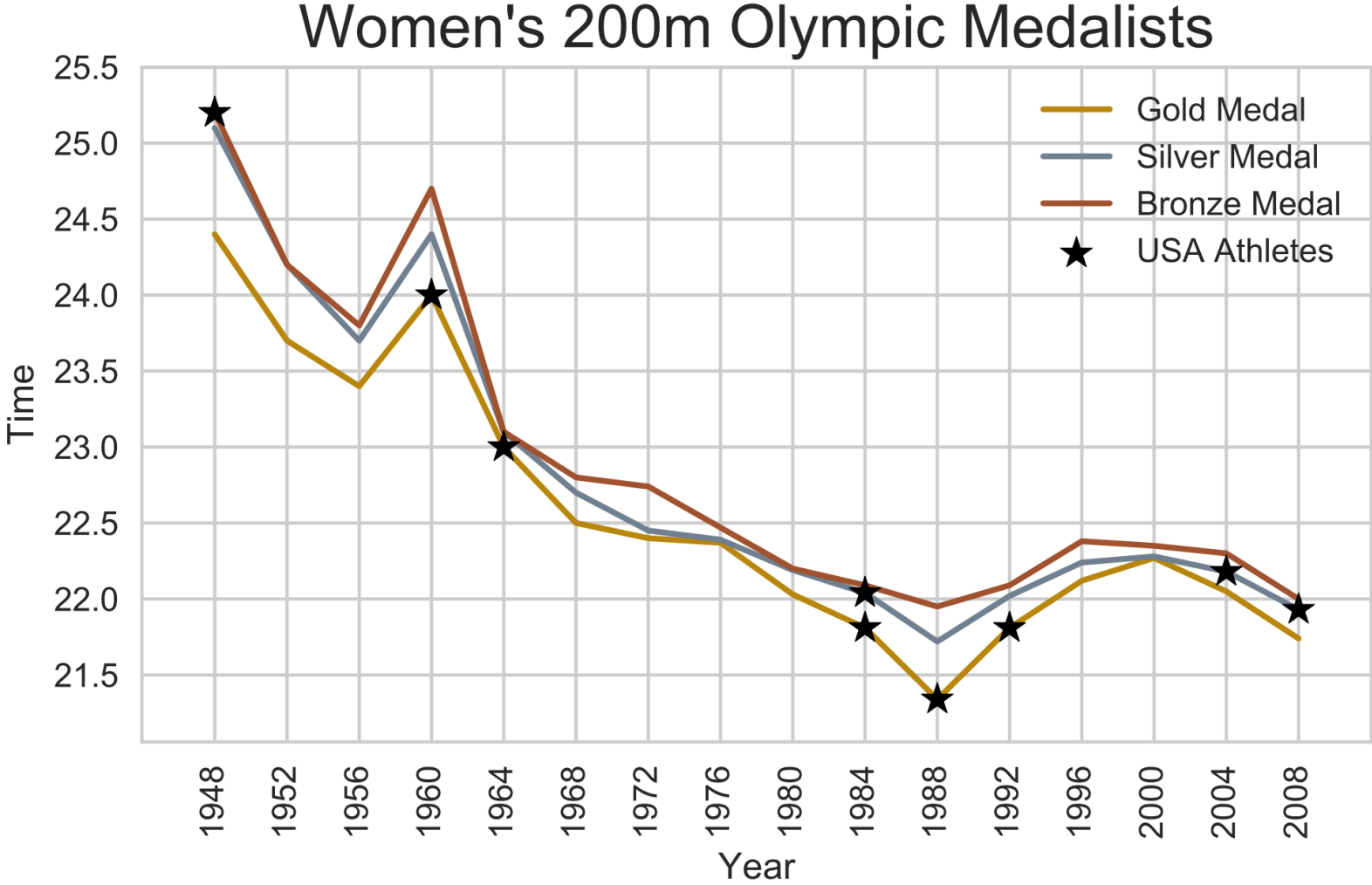
1. Data Visualisation is not Art

	Art	Data Visualisation
Purpose	Evokes emotions and persuades	Communicates data and illuminates patterns
Point of View	Artist's point of view	No point of view
Prescriptiveness	Open to interpretation	Unambiguous
Approach	Free	Prescribed by the data
Rules	Depends, can be bent or broken	Hard and strict

1. Data Visualisation is not Art

	Art	Data Visualisation
Purpose	Evokes emotions and persuades	Communicates data and illuminates patterns
Point of View	Artist's point of view	No point of view
Prescriptiveness	Open to interpretation	Unambiguous
Approach	Free	Prescribed by the data
Rules	Depends, can be bent or broken	Hard and strict
Creativity	Encouraged, vital	Possible, but less important

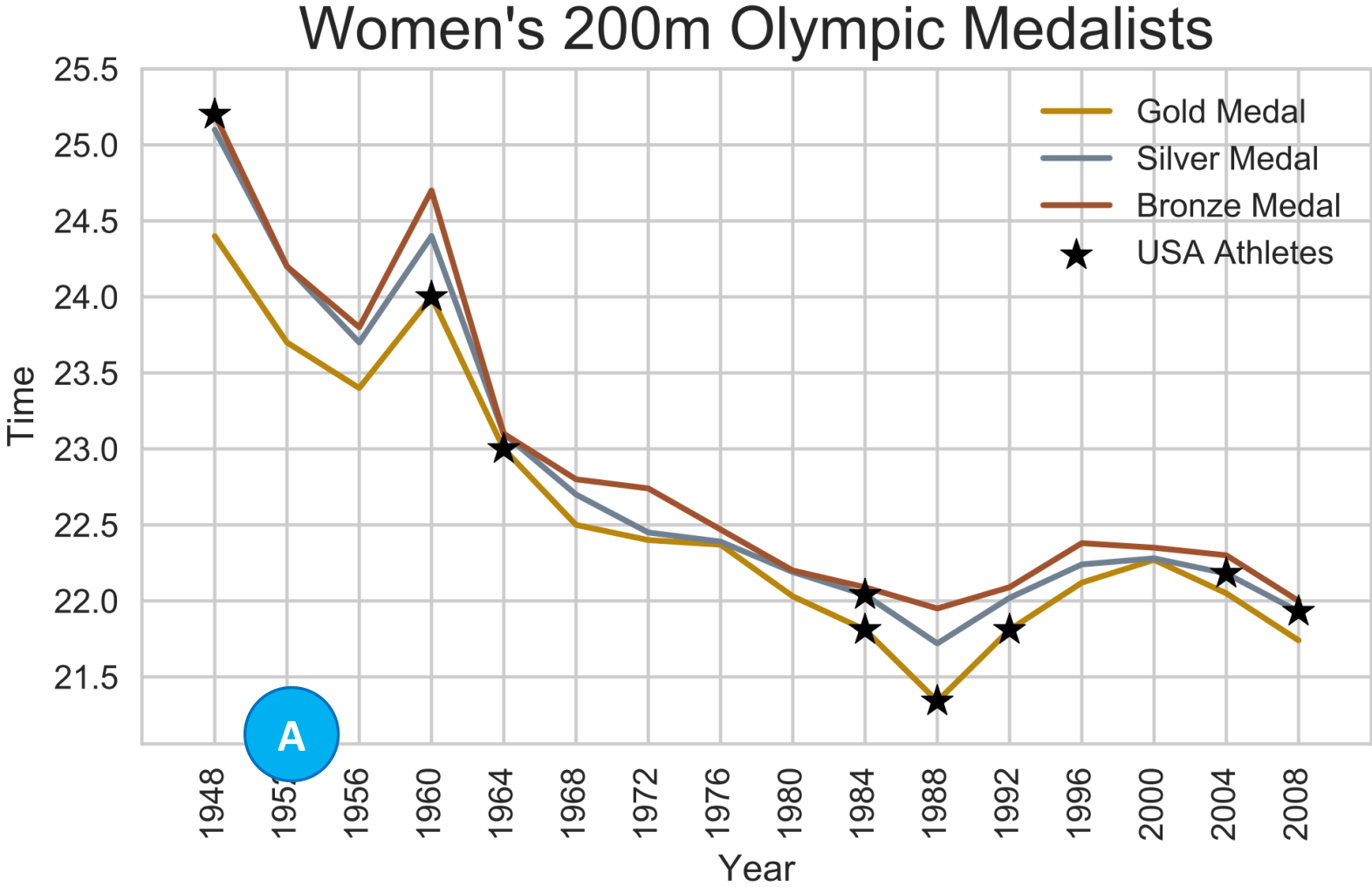
1. Anatomy of a Chart



1. Anatomy of a Chart



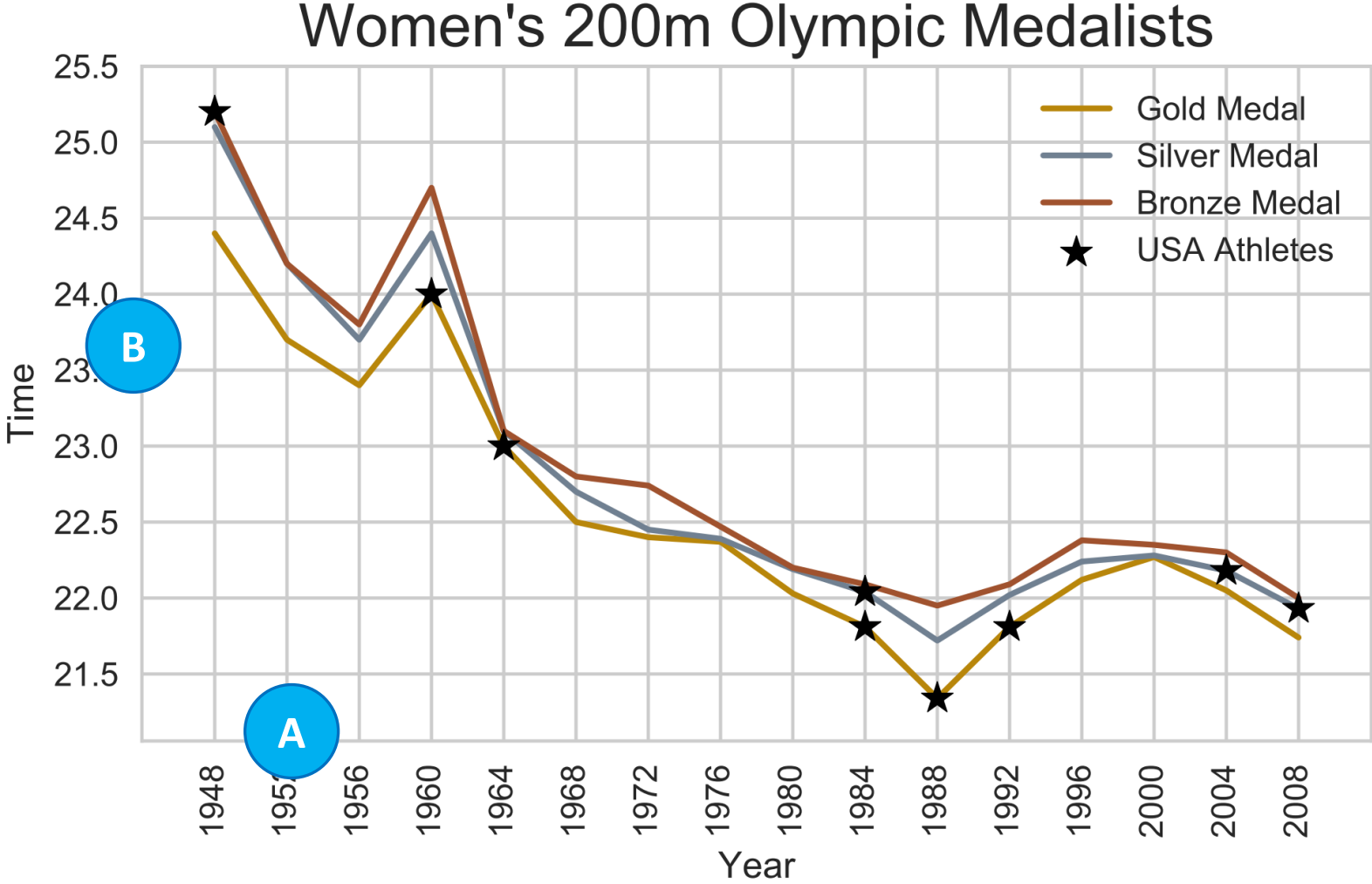
x-axis



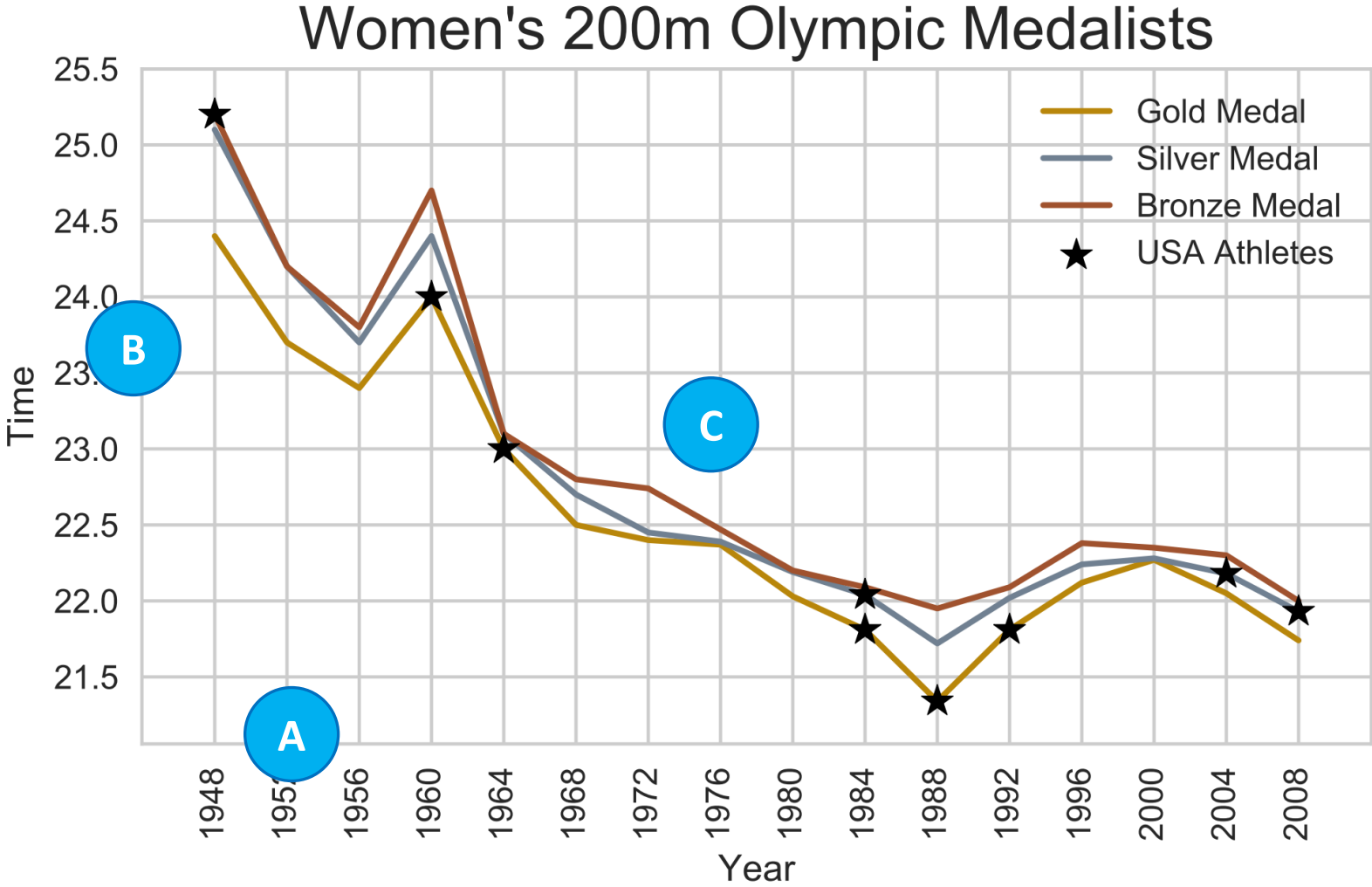
1. Anatomy of a Chart

A *x-axis*

B *y-axis*



1. Anatomy of a Chart



x-axis

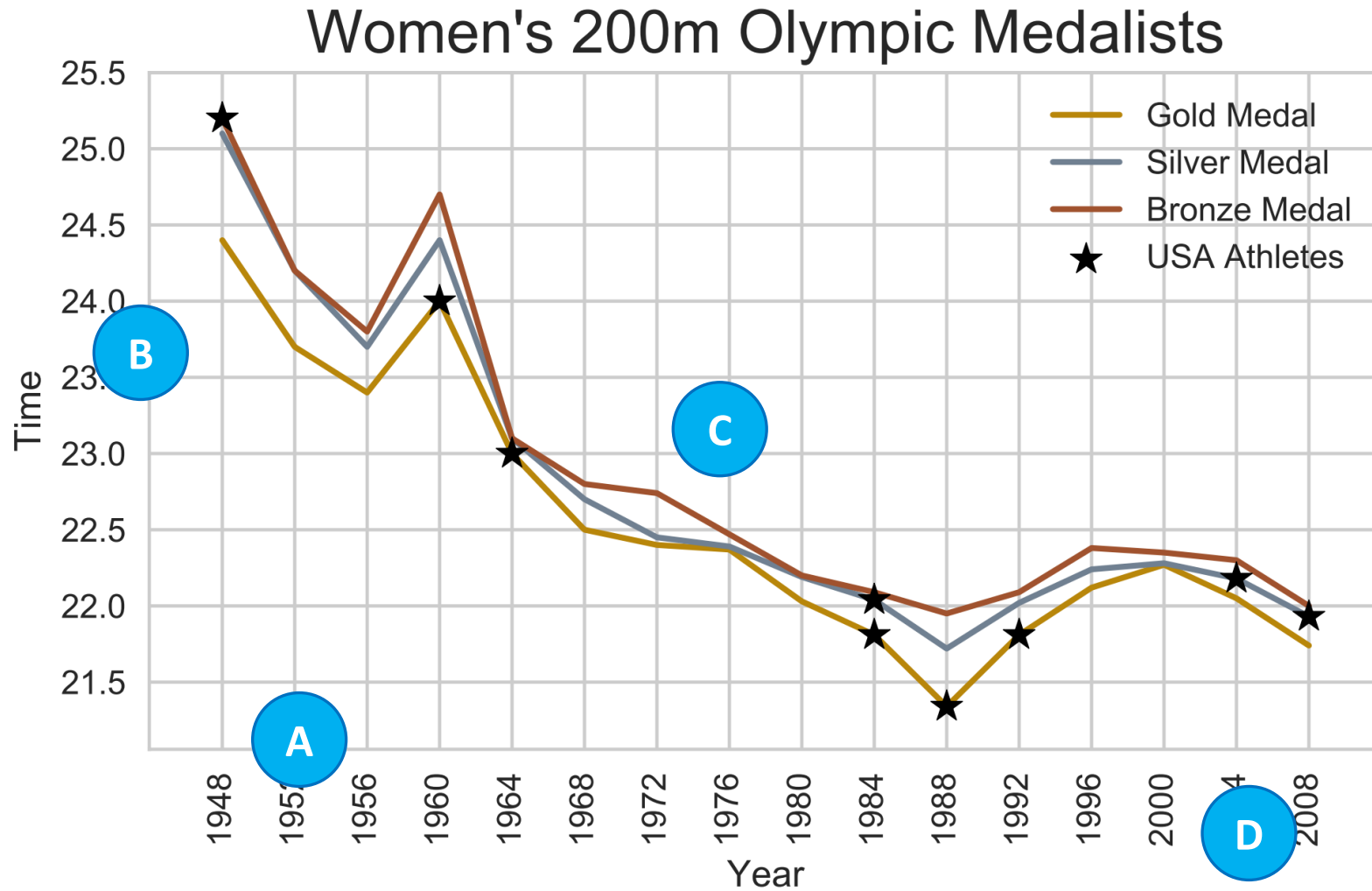


y-axis



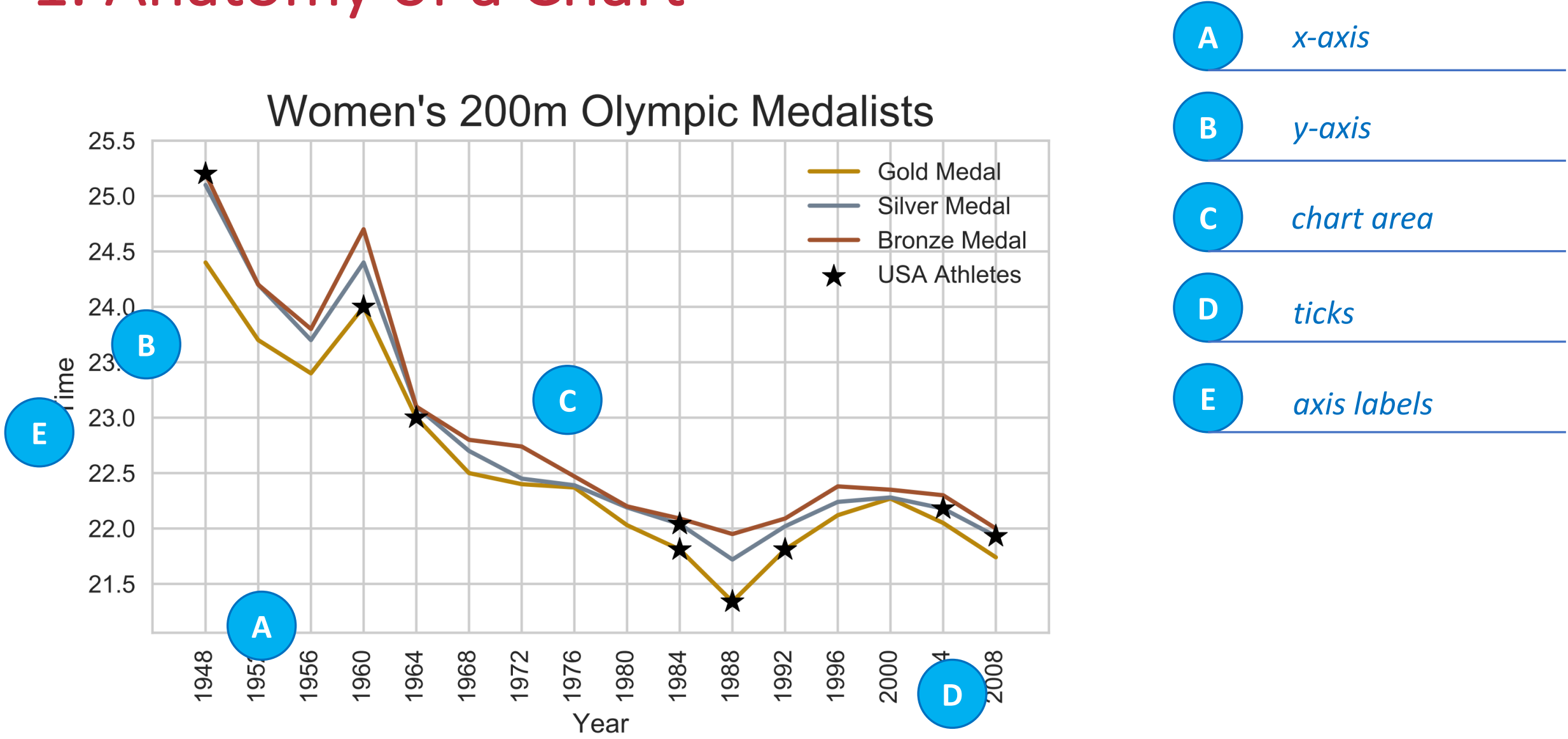
chart area

1. Anatomy of a Chart

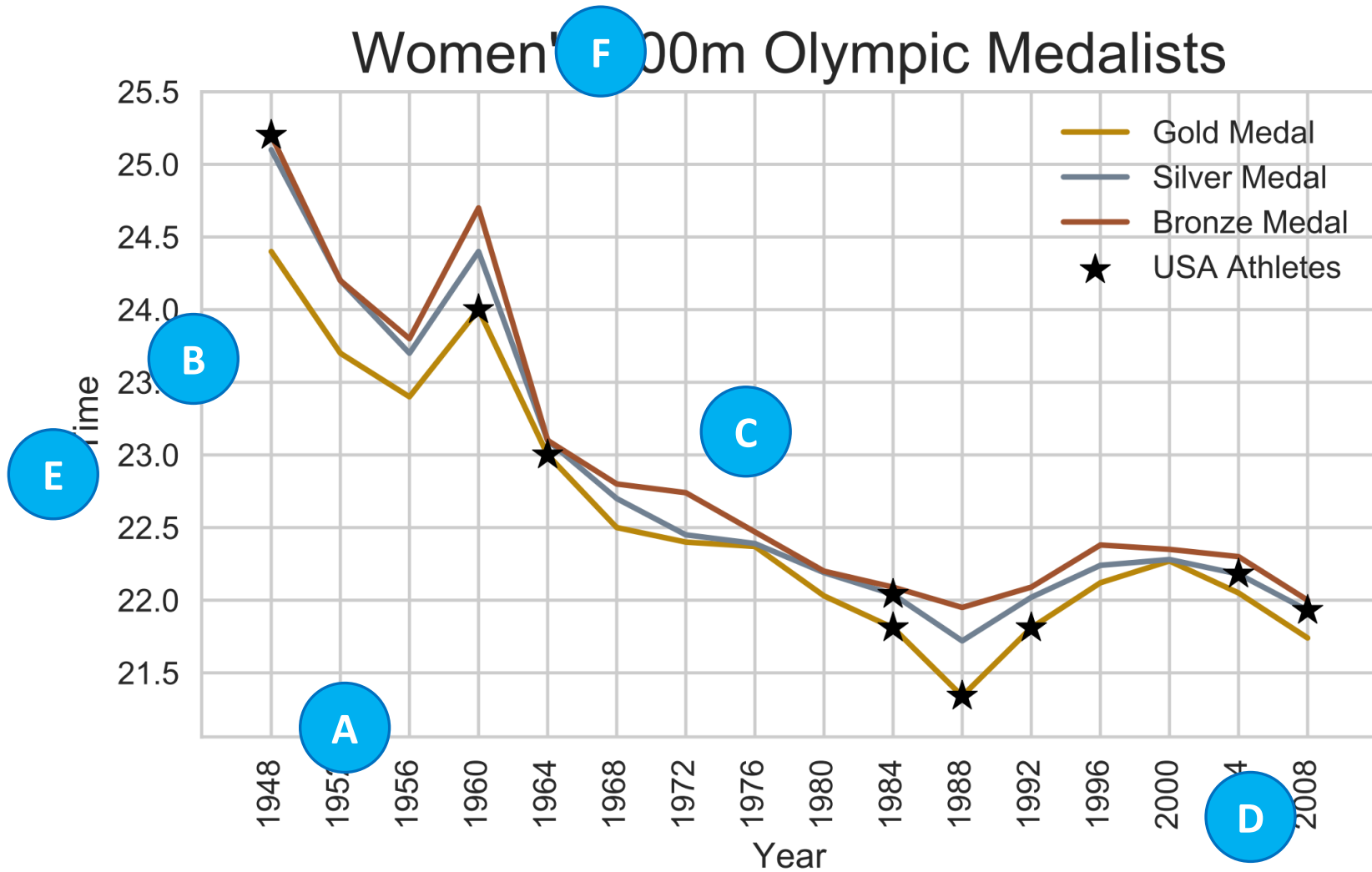


- A *x-axis*
- B *y-axis*
- C *chart area*
- D *ticks*

1. Anatomy of a Chart

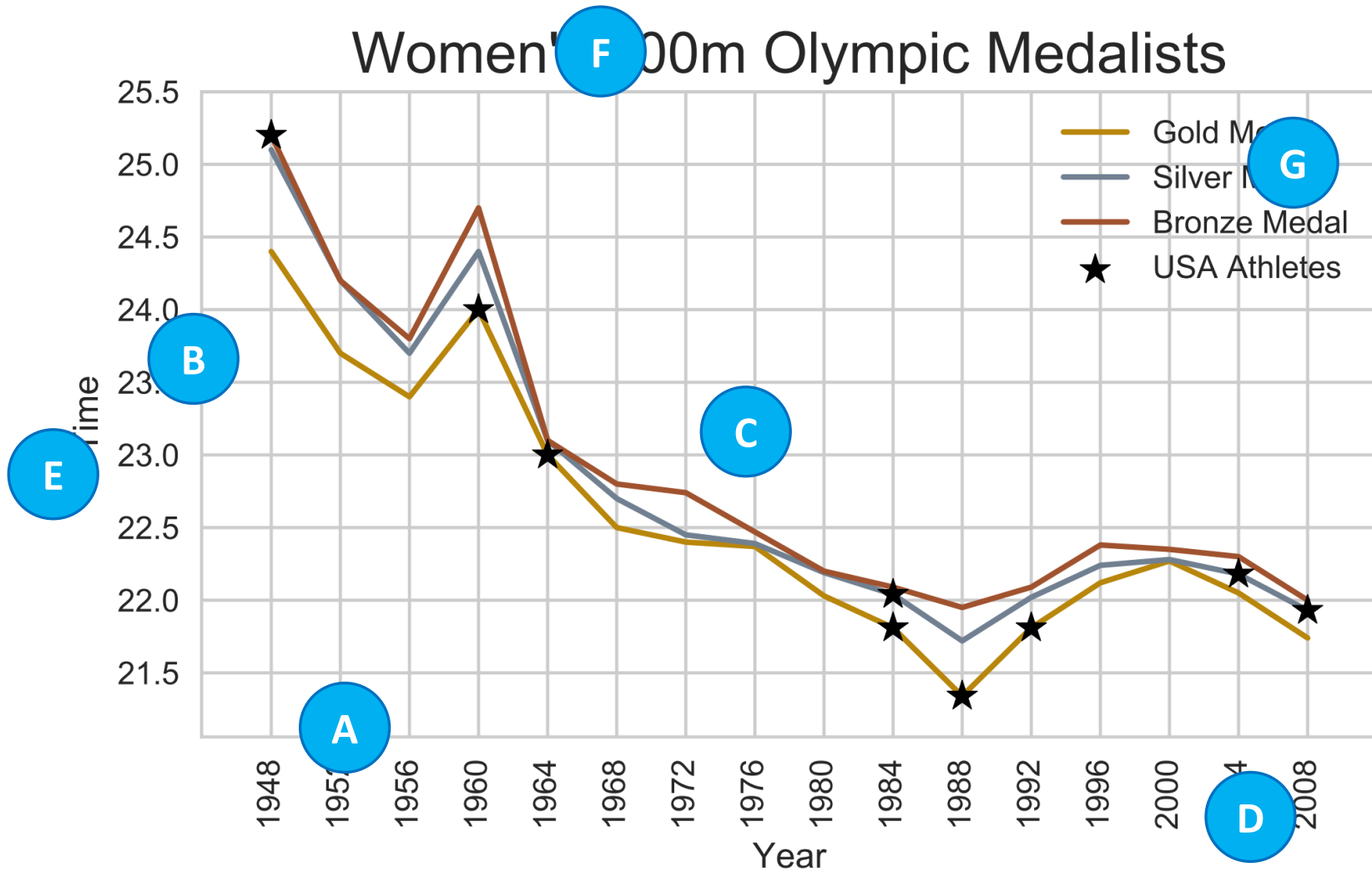


1. Anatomy of a Chart



- A *x-axis*
- B *y-axis*
- C *chart area*
- D *ticks*
- E *axis labels*
- F *title*

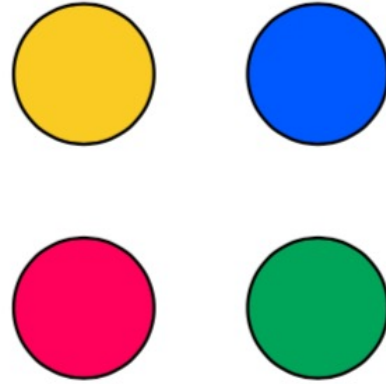
1. Anatomy of a Chart



- A** *x-axis*
- B** *y-axis*
- C** *chart area*
- D** *ticks*
- E** *axis labels*
- F** *title*
- G** *key / legend*

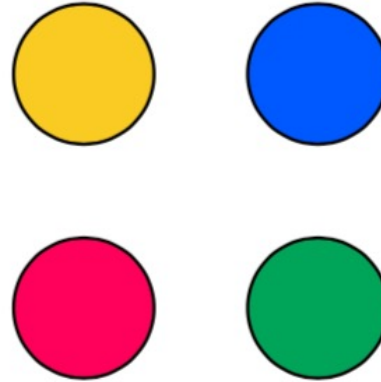
1. Data Types

1. Data Types

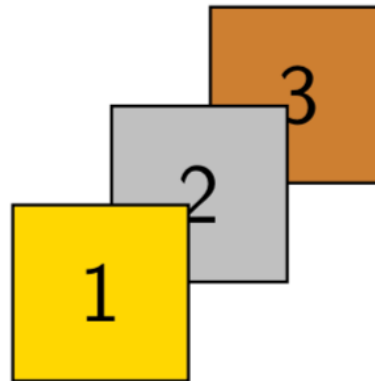


Nominal

1. Data Types

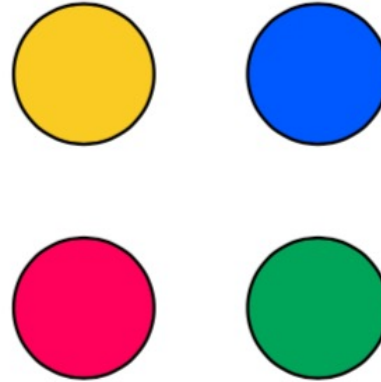


Nominal

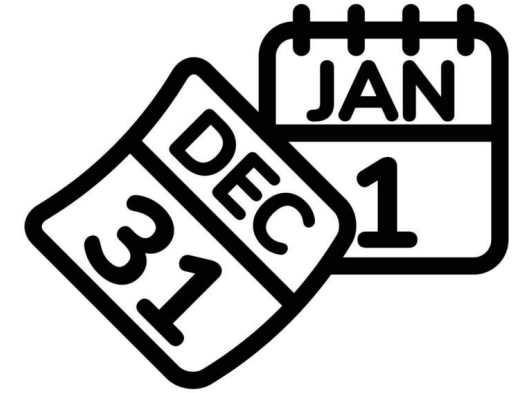


Ordinal

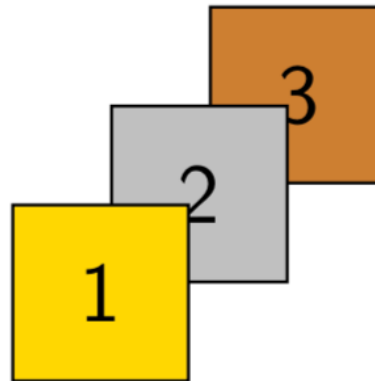
1. Data Types



Nominal

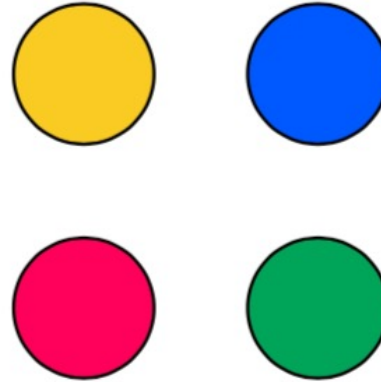


Interval

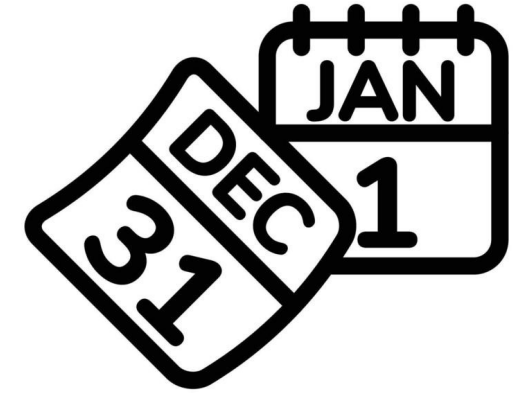


Ordinal

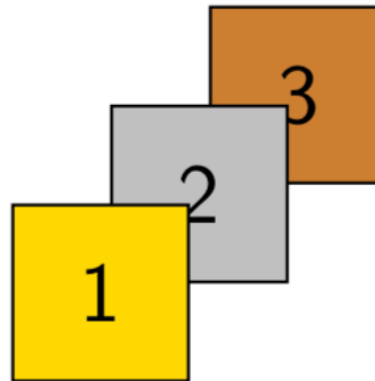
1. Data Types



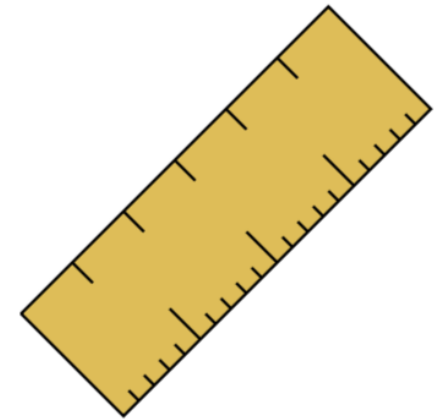
Nominal



Interval



Ordinal



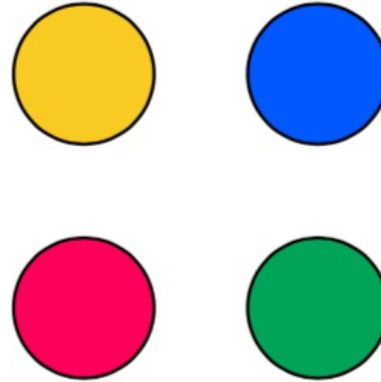
Ratio

1. Data Types

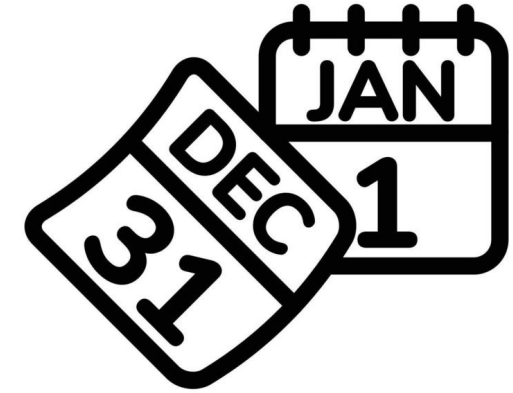
<i>DATA TYPE</i>	<i>ORDER</i>	<i>DISTANCE</i>	<i>ABSOLUTE ZERO</i>
<i>Nominal</i>	✗	✗	✗
<i>Ordinal</i>	✓	✗	✗
<i>Interval</i>	✓	✓	✗
<i>Ratio</i>	✓	✓	✓

1. Data Types

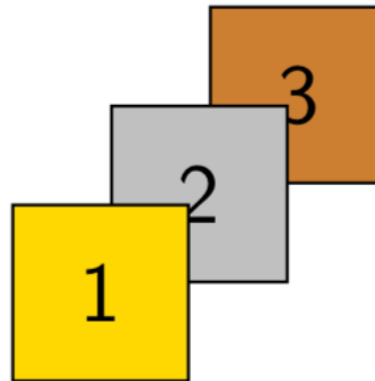
- Weight of bananas in kilograms
- Date of Birth
- Age
- Likert Scale
(strongly agree, neutral, disagree, strongly disagree)
- Favourite colour
- Temperature in Celcius
- Temperature in Kelvin
- Sex



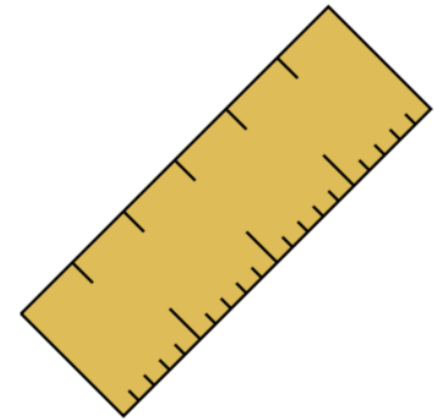
Nominal



Interval



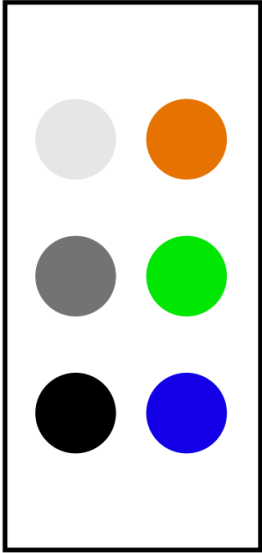
Ordinal



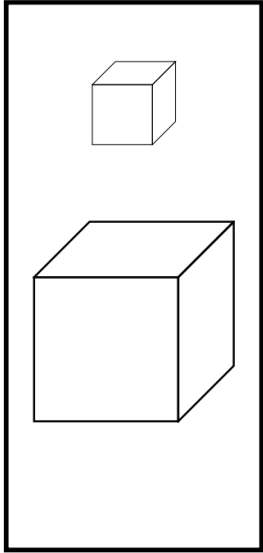
Ratio

1. Elements Available

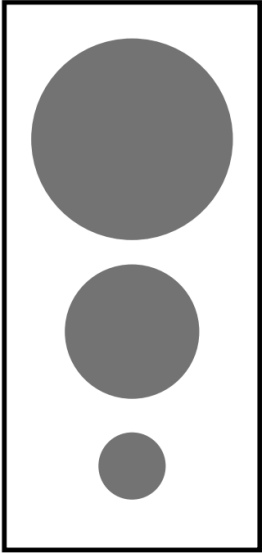
Density/
Colour



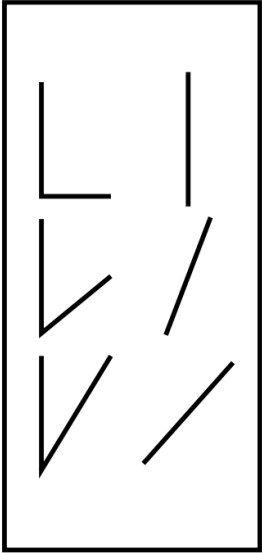
Volume



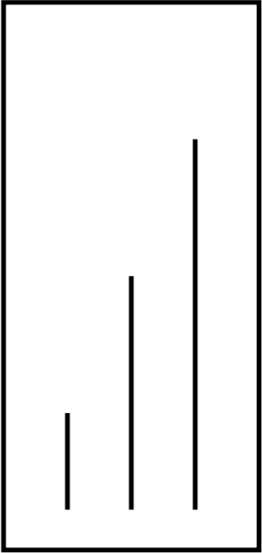
Area



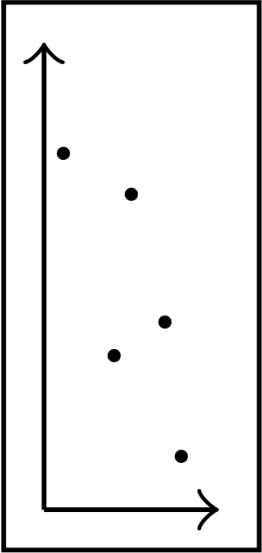
Angle/
Slope



Length

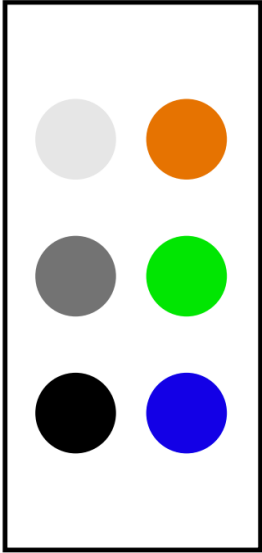


Position

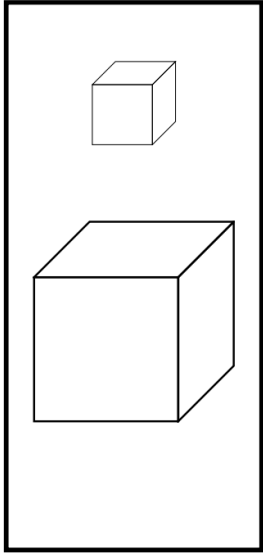


1. Elements Available

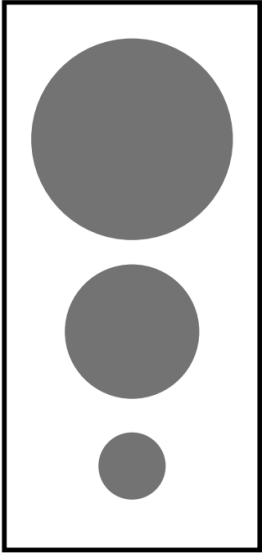
Density/
Colour



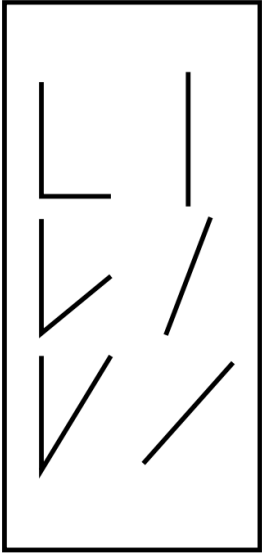
Volume



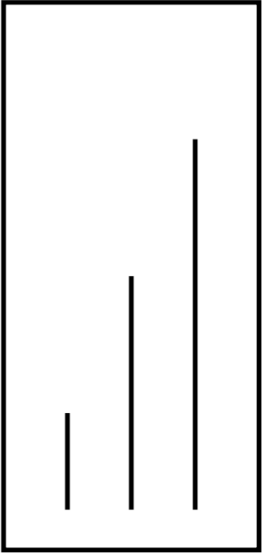
Area



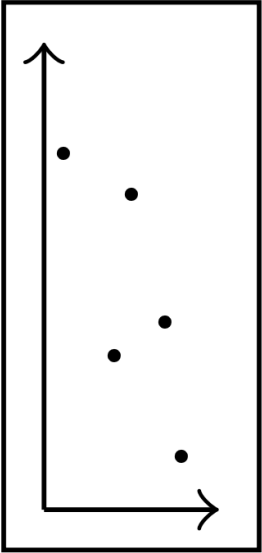
Angle/
Slope



Length



Position

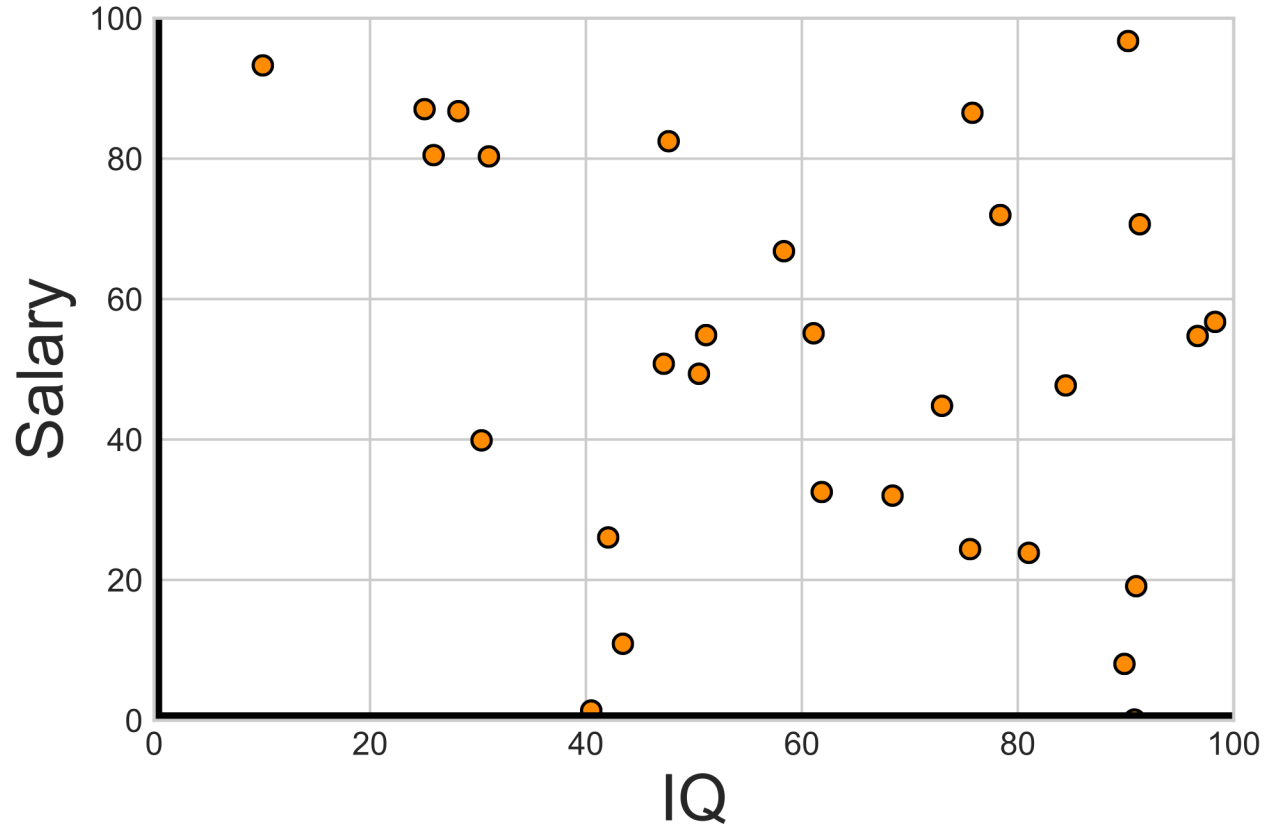


Less Accurate

More Accurate

2) Comparing 2 or More Variables

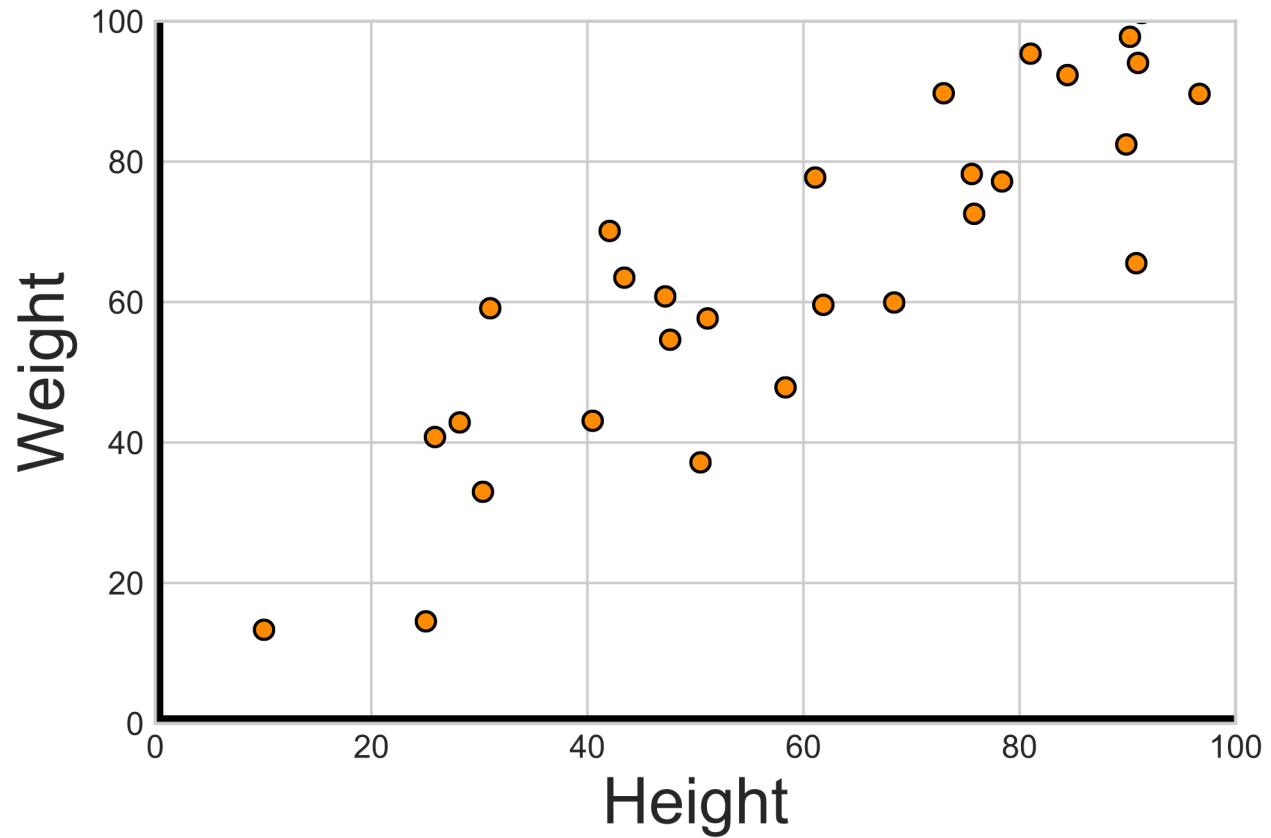
2. Scatter Plots



ratio vs ratio

no relationship

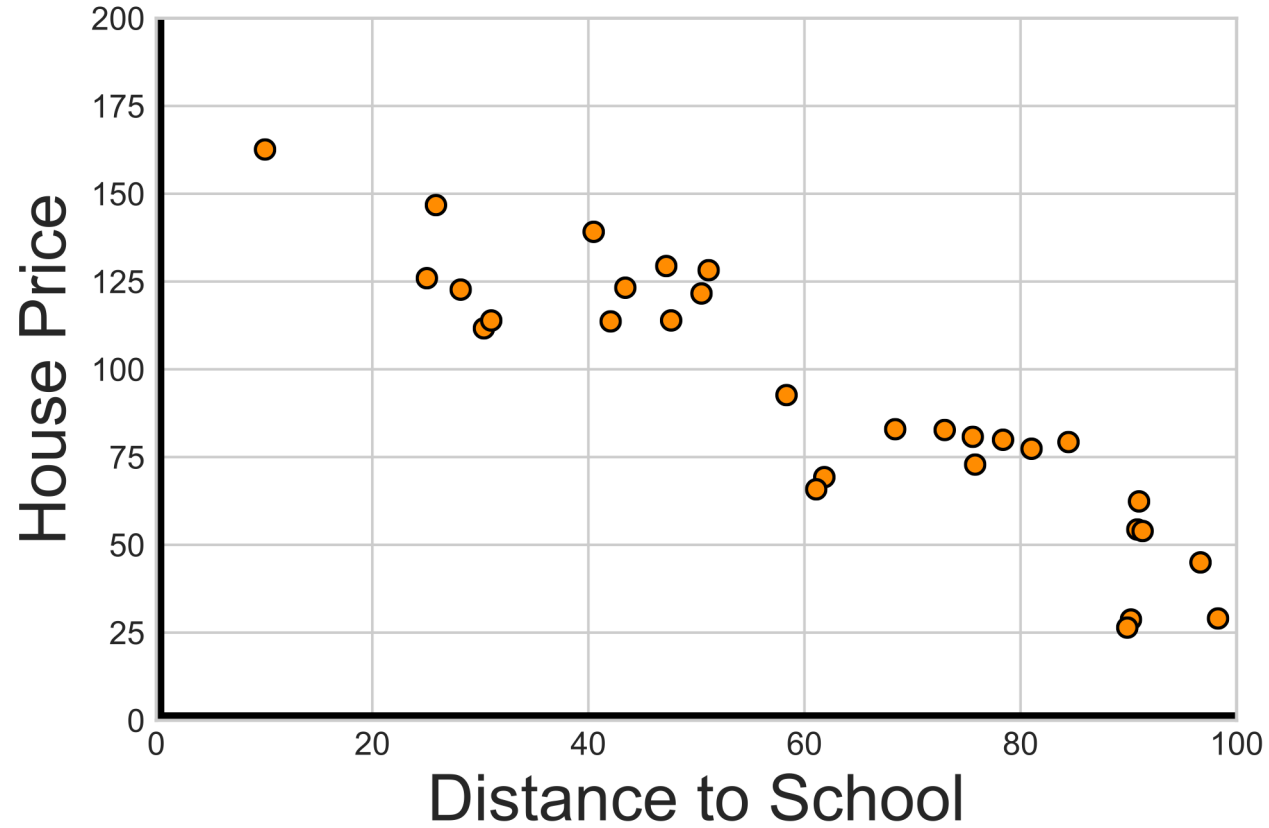
2. Scatter Plots



ratio vs ratio

positive relationship

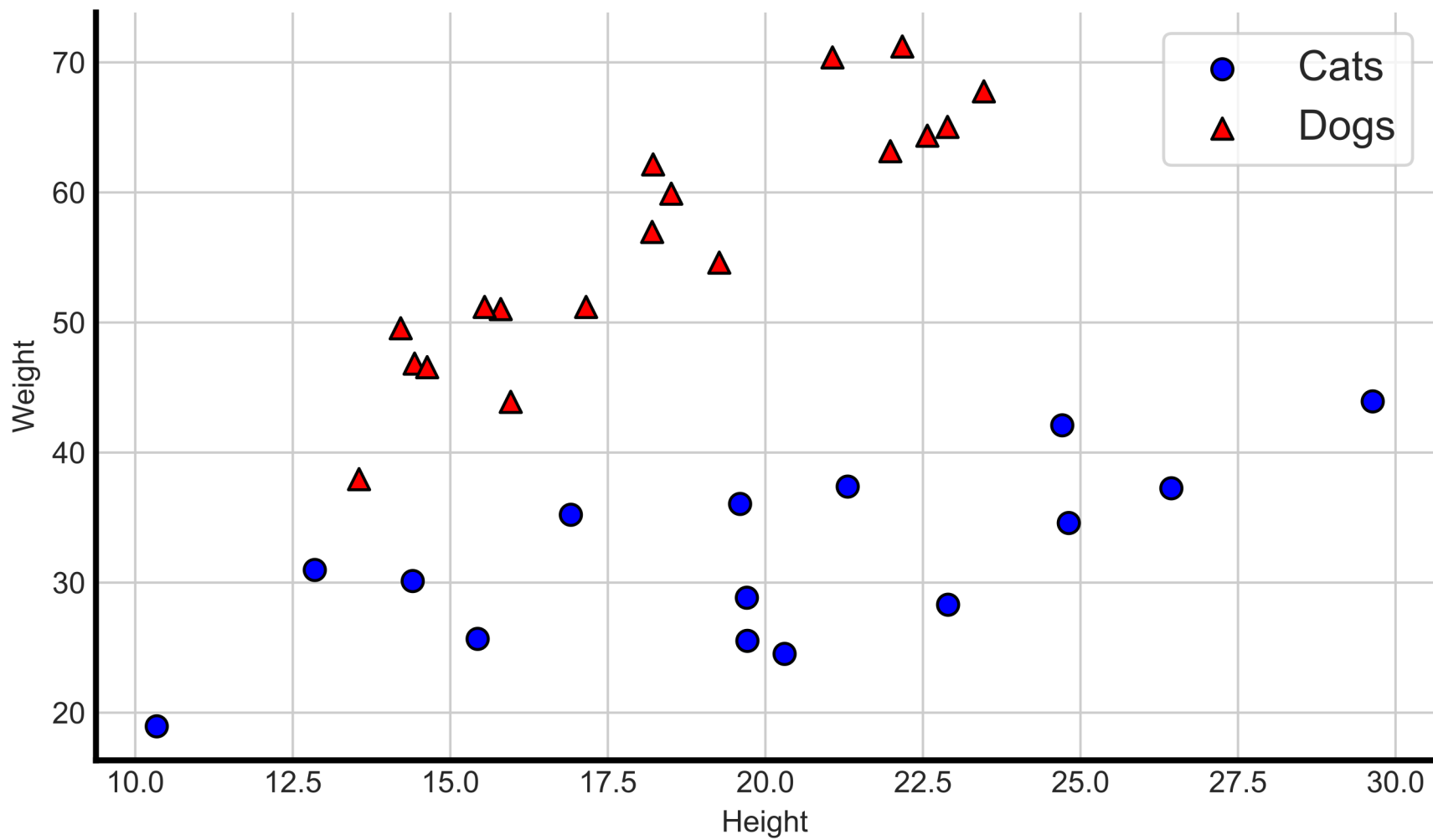
2. Scatter Plots



ratio vs ratio

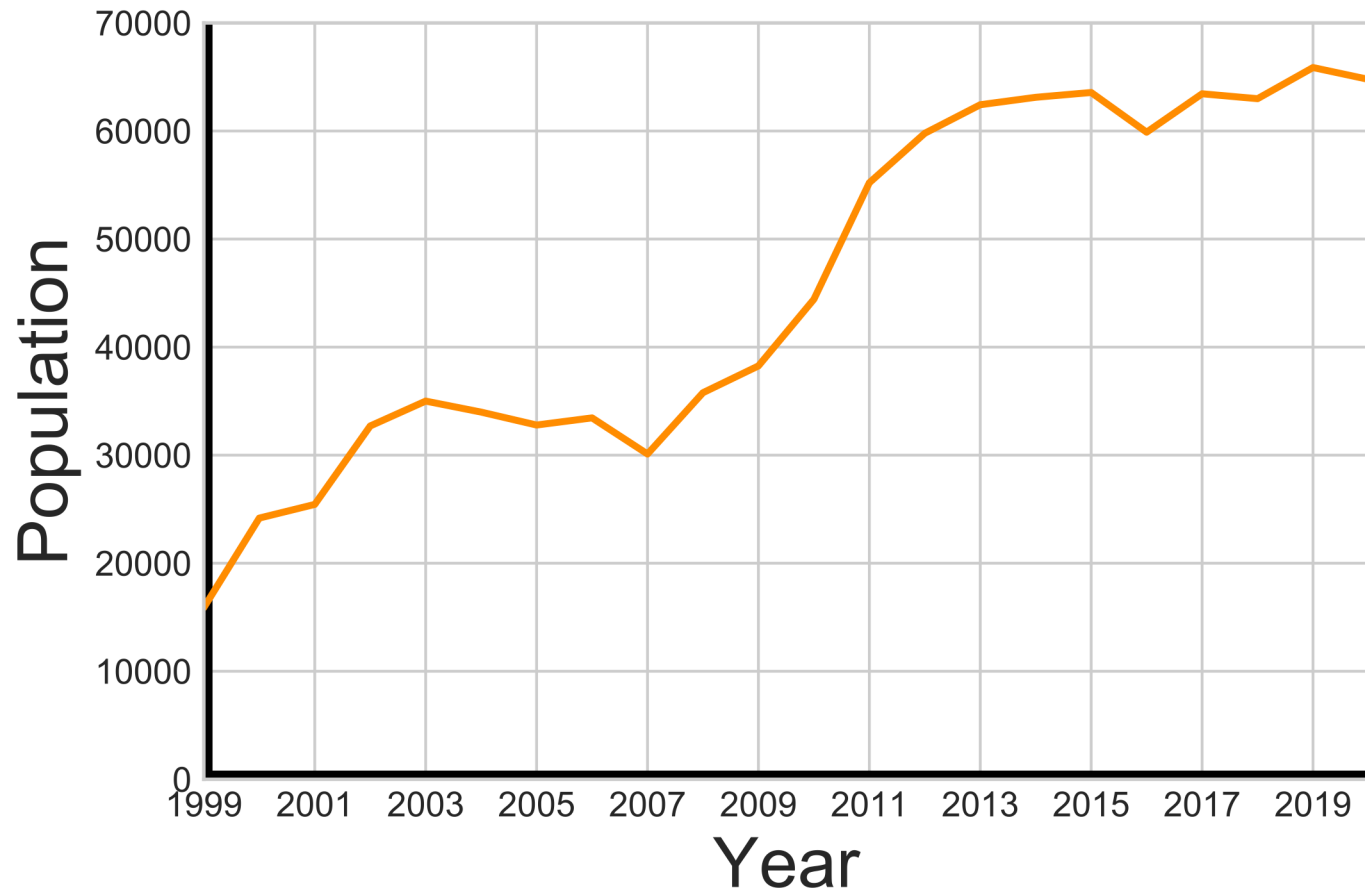
negative relationship

2. Scatter Plots



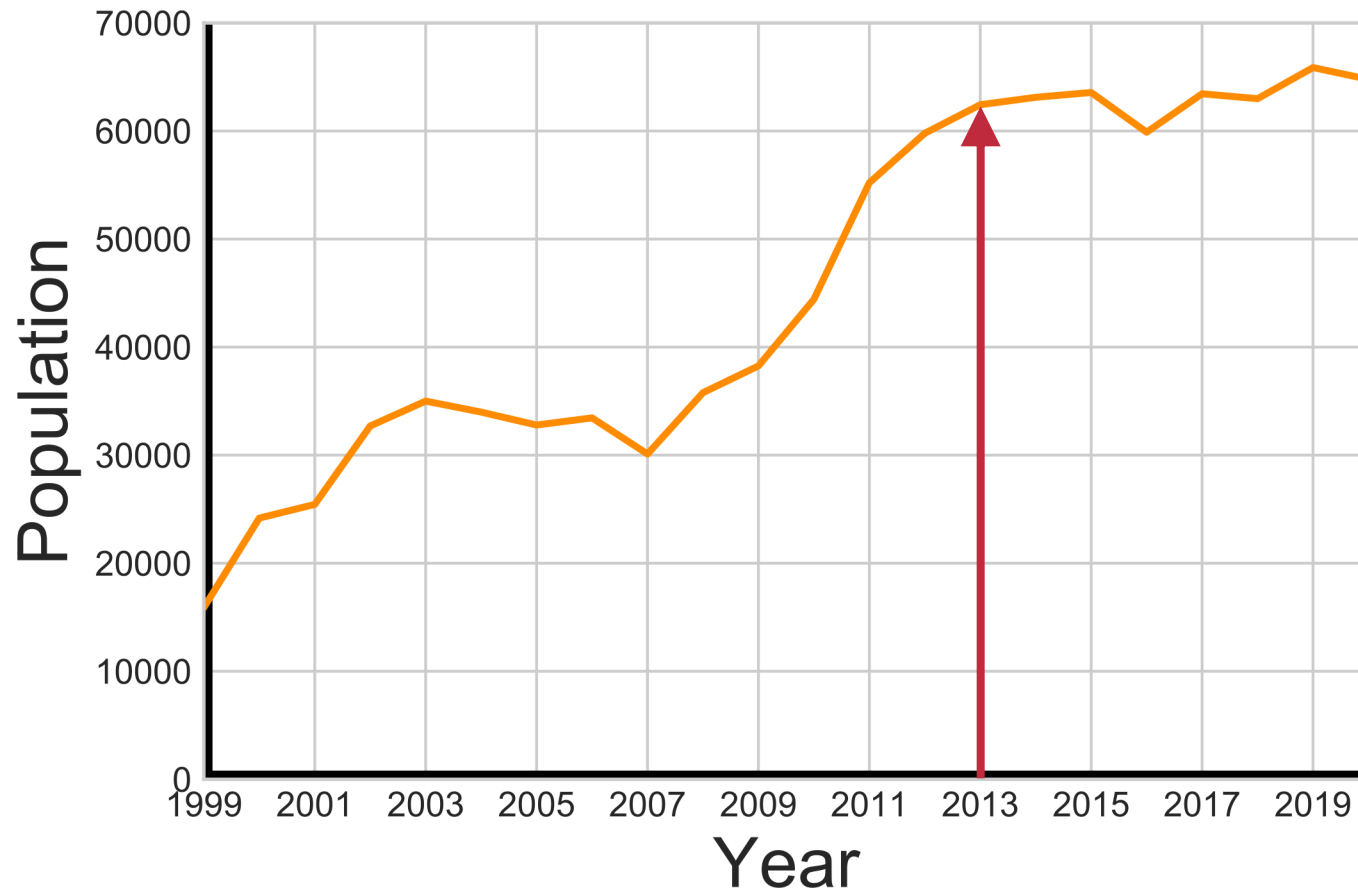
2. Line Plots

ratio vs ratio / interval



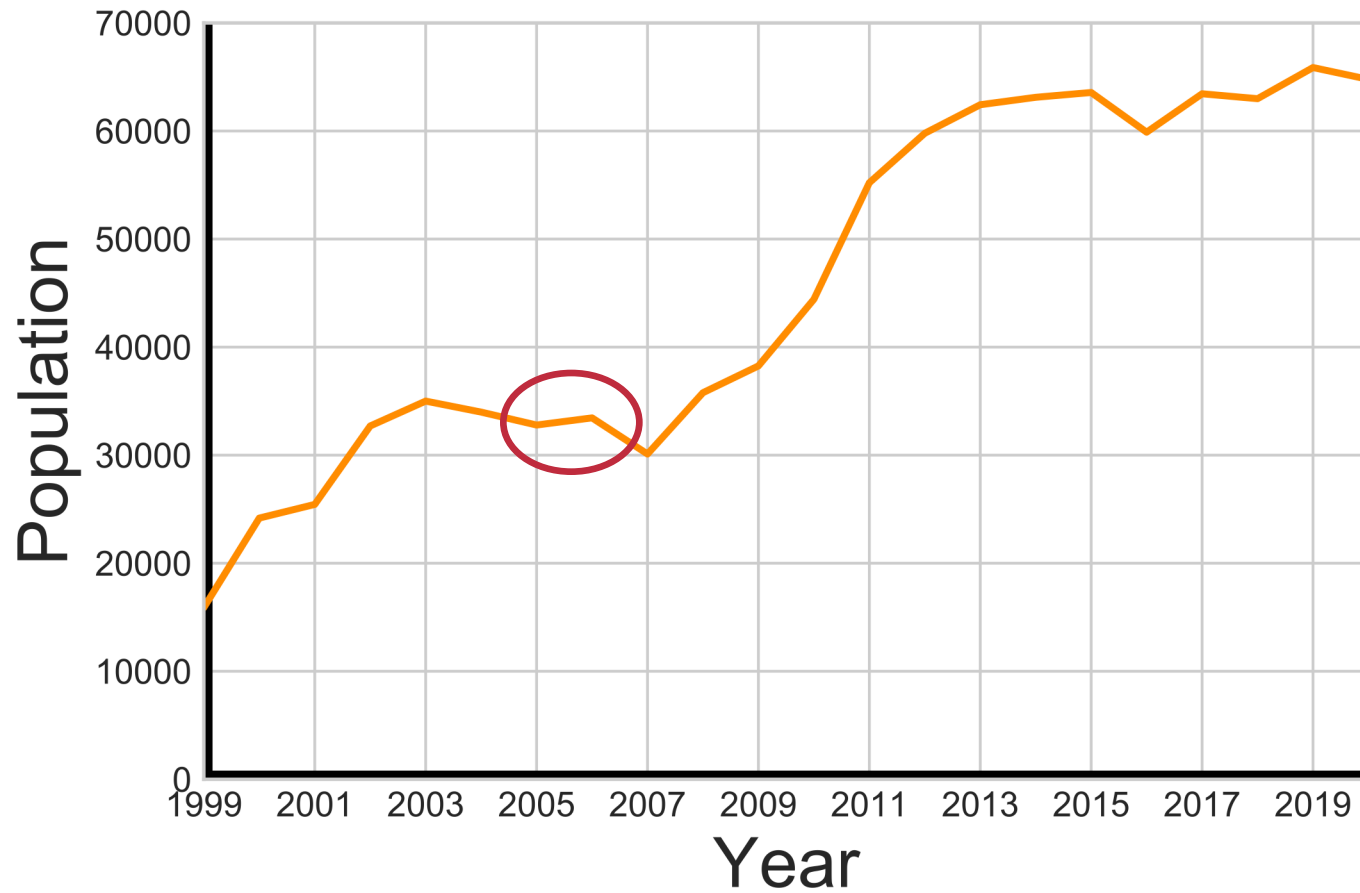
2. Line Plots

ratio vs ratio / interval



1 to 1

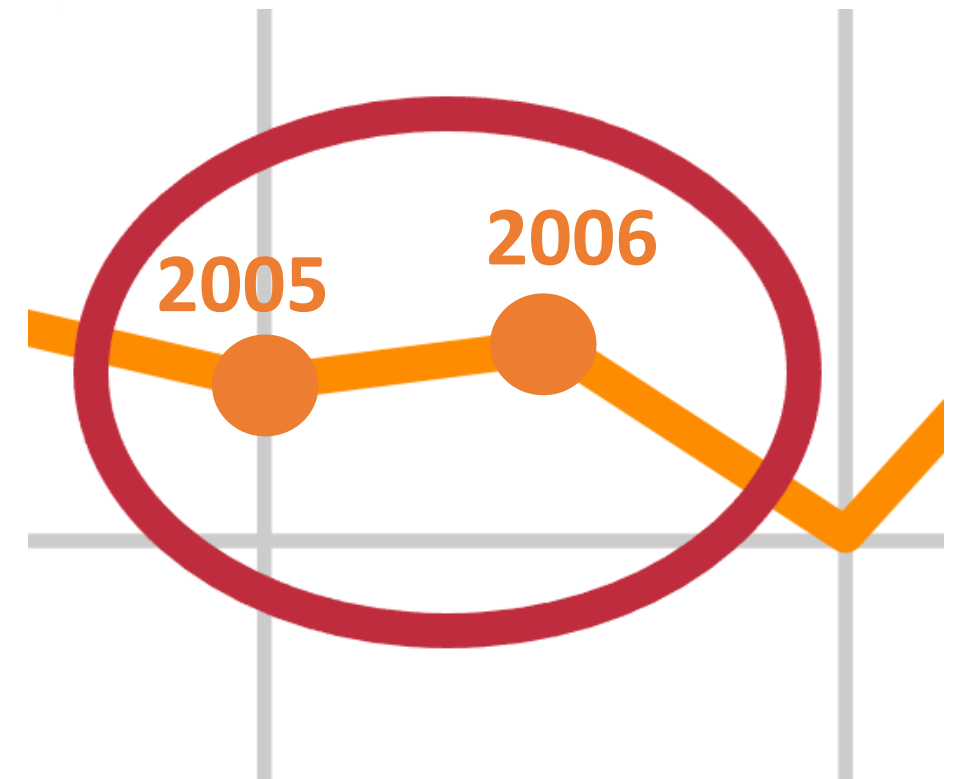
2. Line Plots



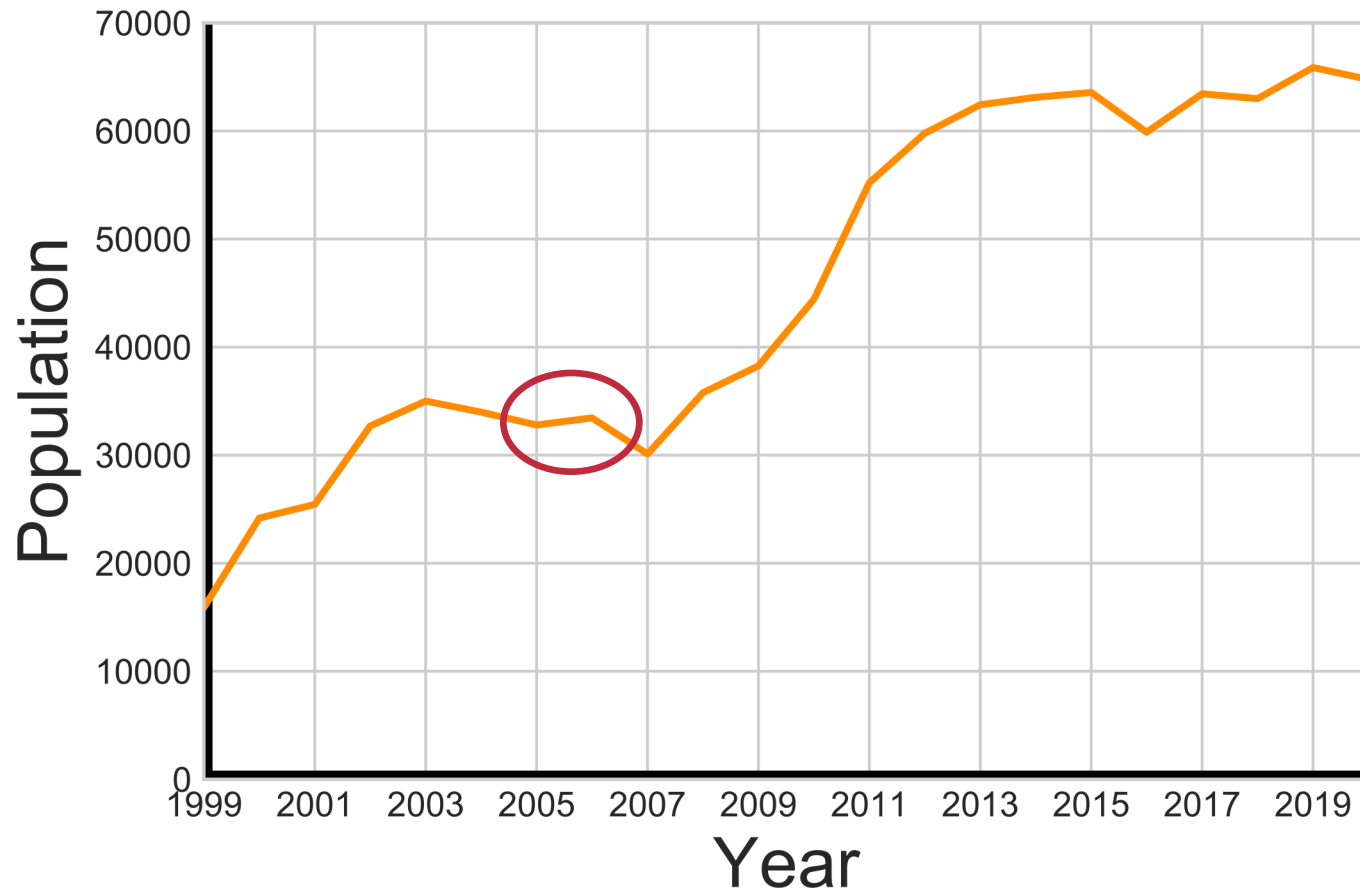
ratio vs ratio / interval

1 to 1

can interpolate



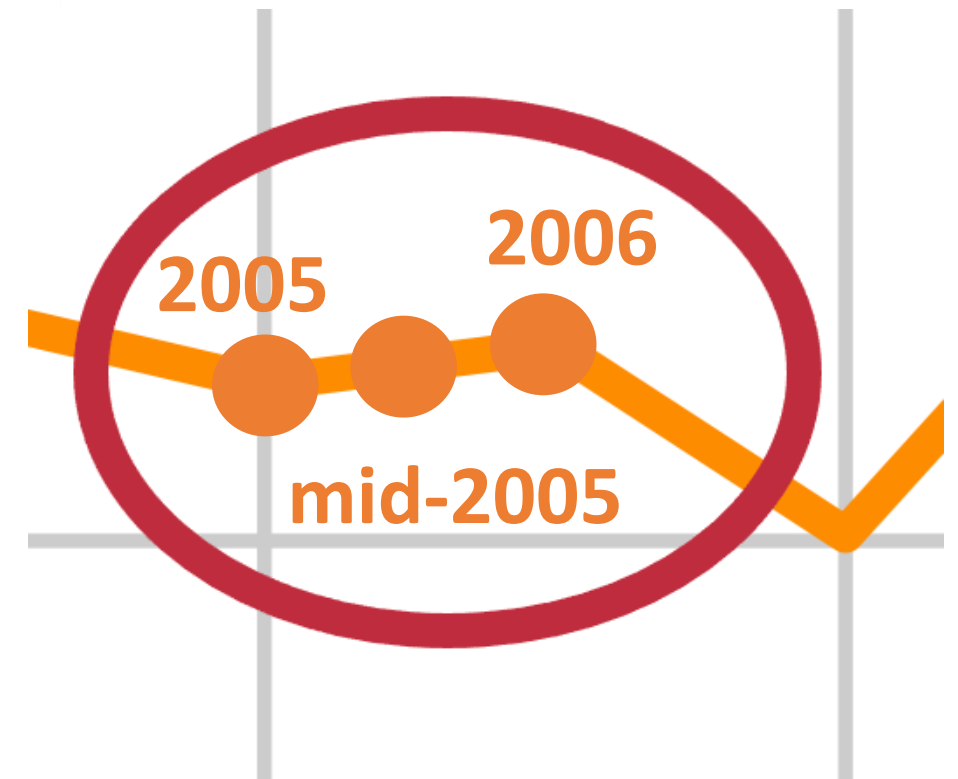
2. Line Plots



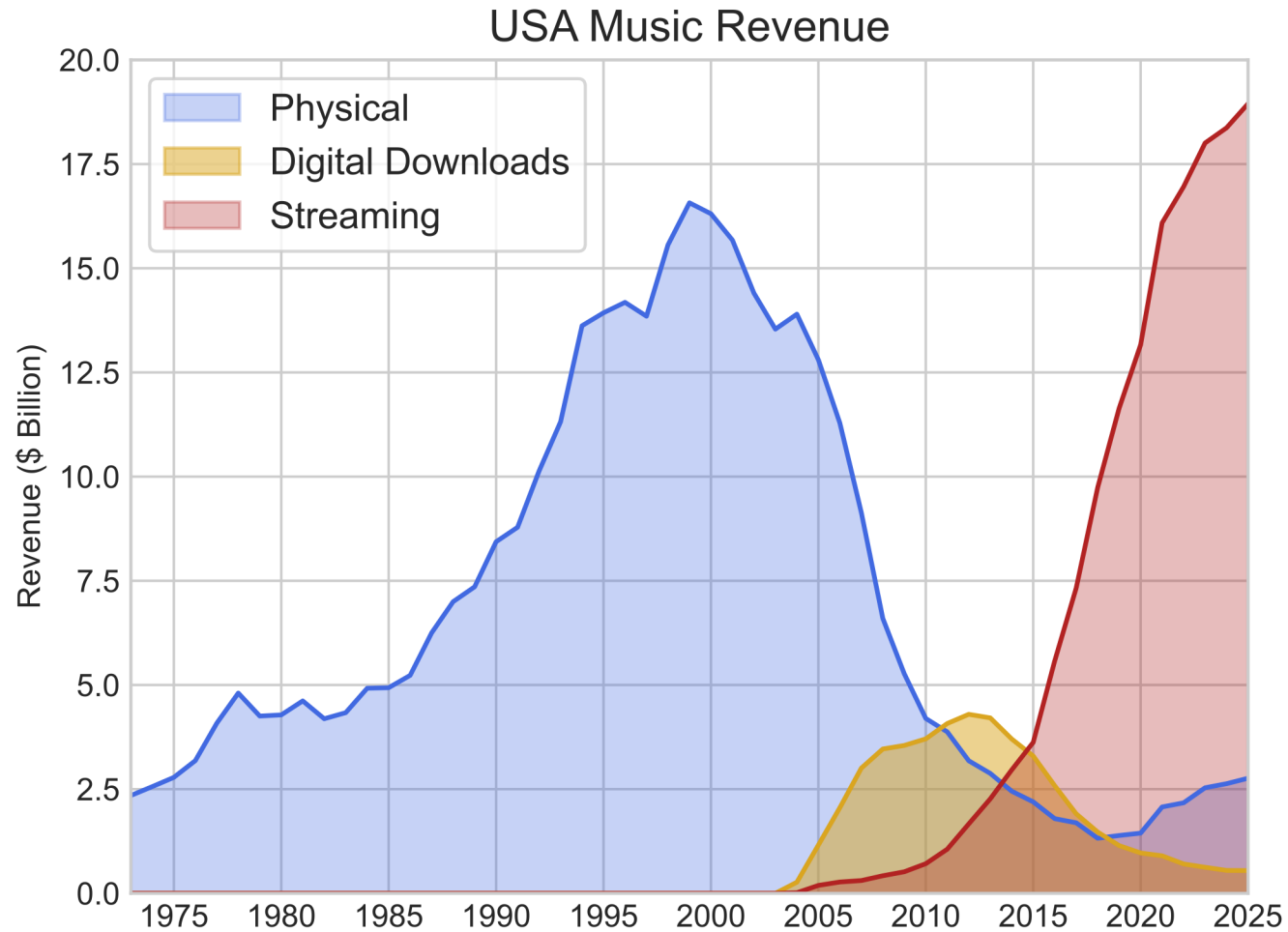
ratio vs ratio / interval

1 to 1

can interpolate



2. Area Plots

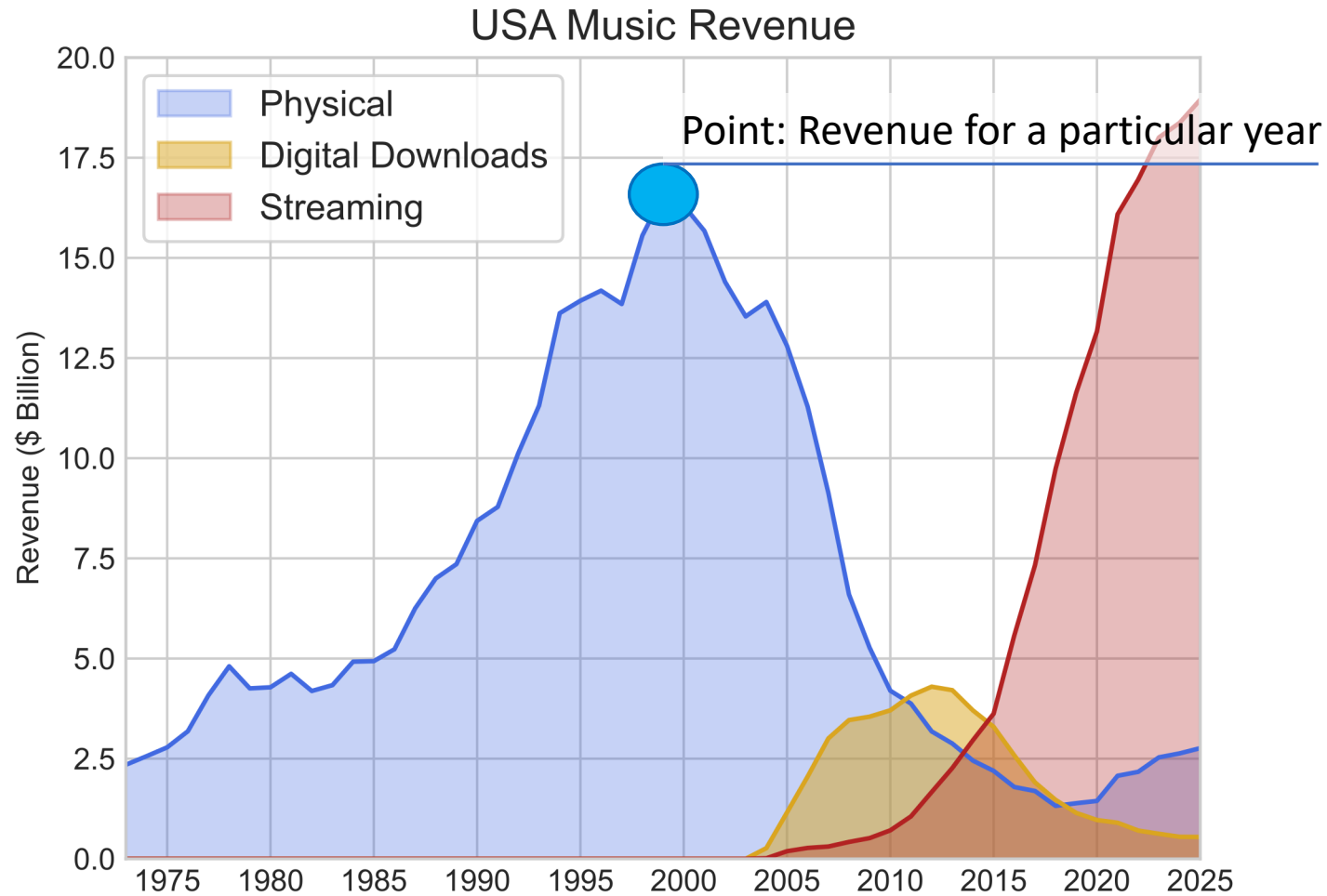


ratio vs ratio / interval

1 to 1

can interpolate

2. Area Plots



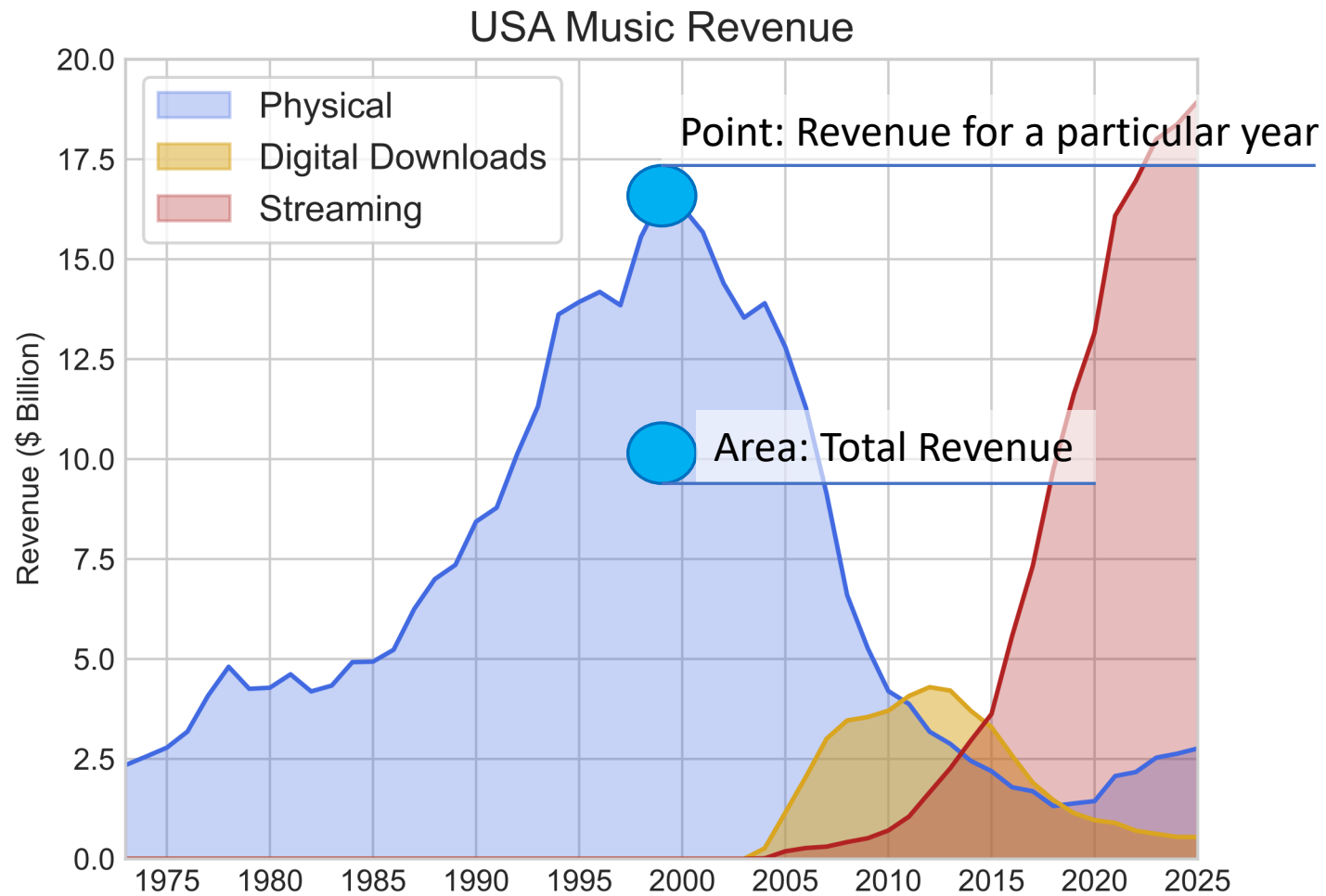
ratio vs ratio / interval

1 to 1

can interpolate

*both value and integral
are meaningful*

2. Area Plots



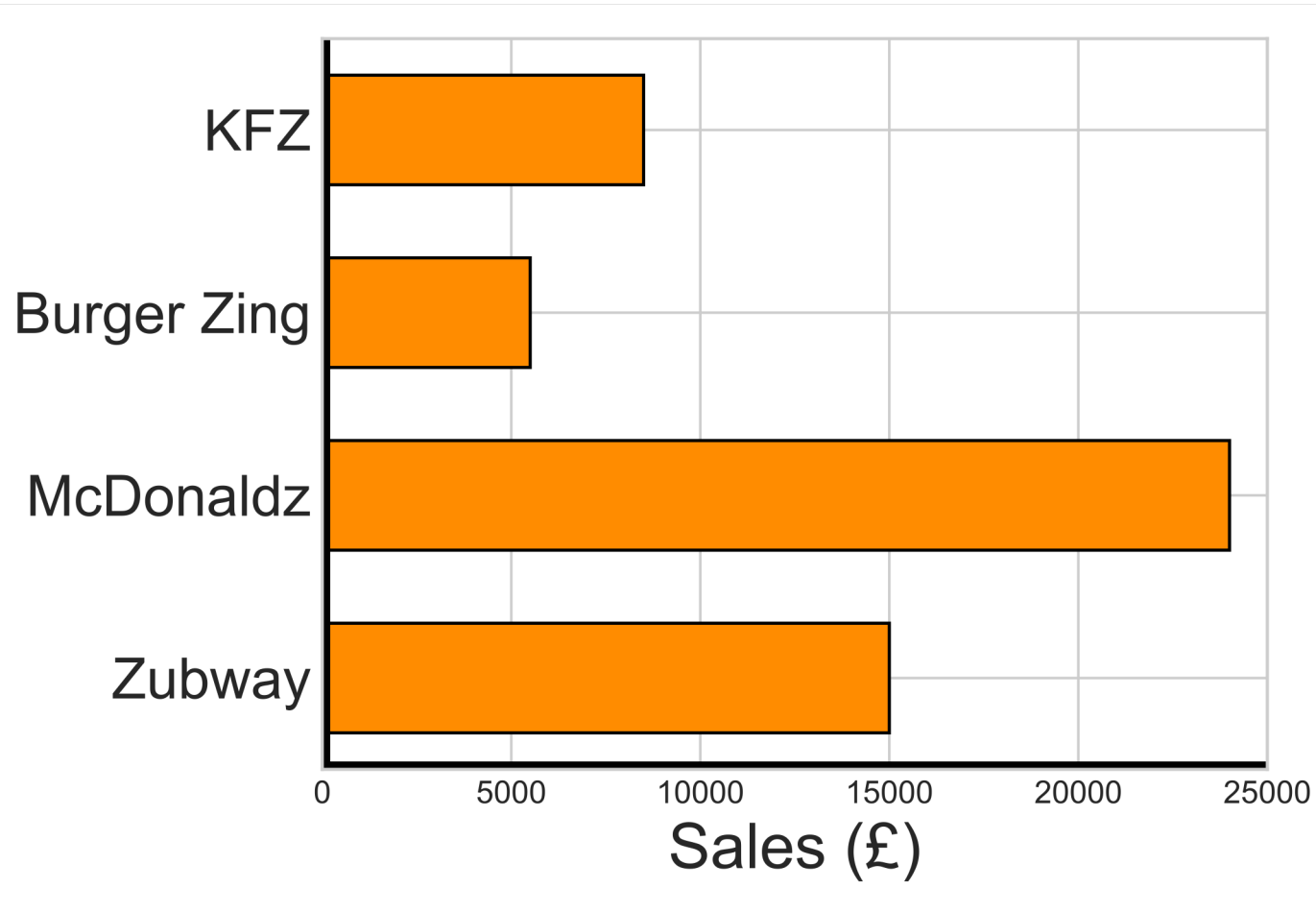
ratio vs ratio / interval

1 to 1

can interpolate

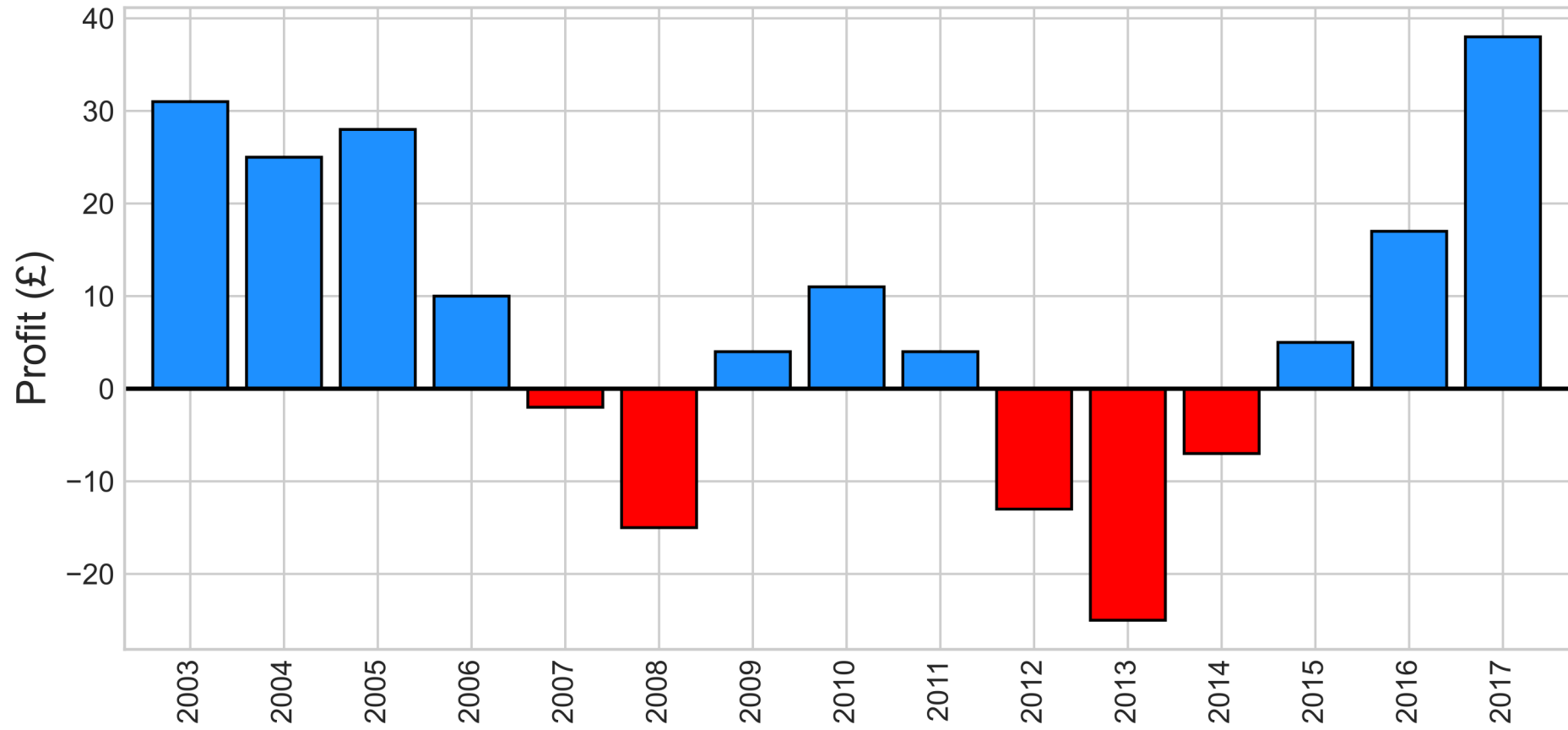
*both value and integral
are meaningful*

2. Bar Charts

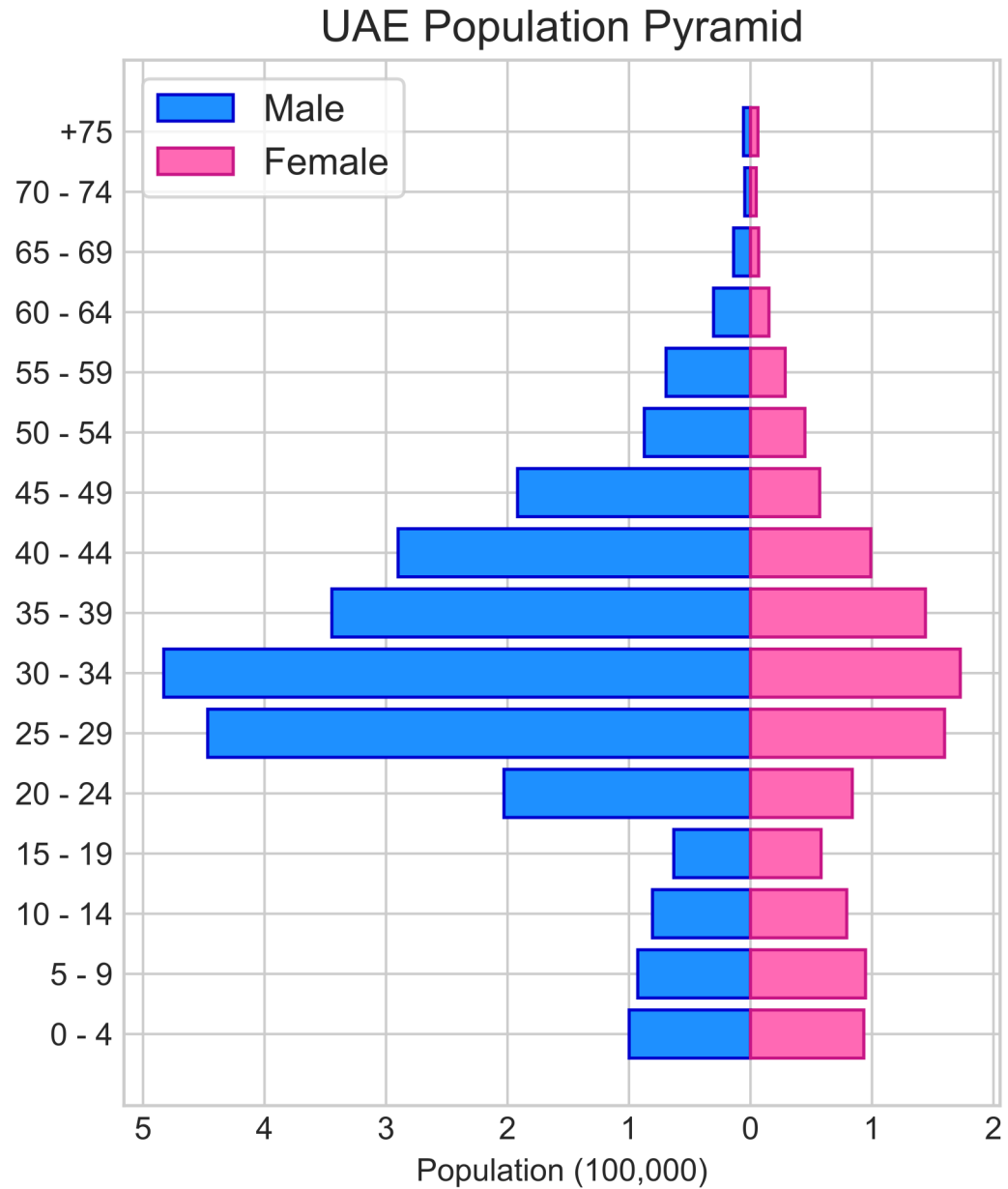


nominal/ordinal
VS
ratio

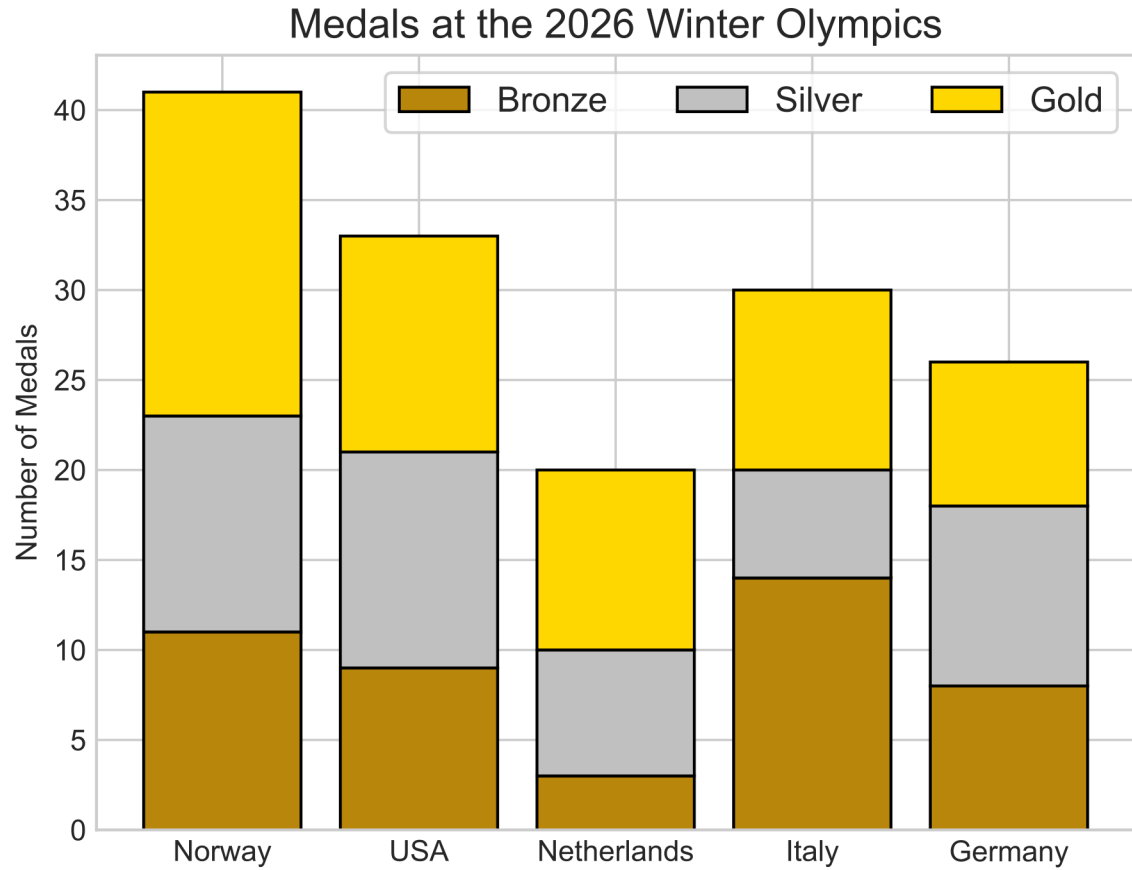
2. Bar Charts



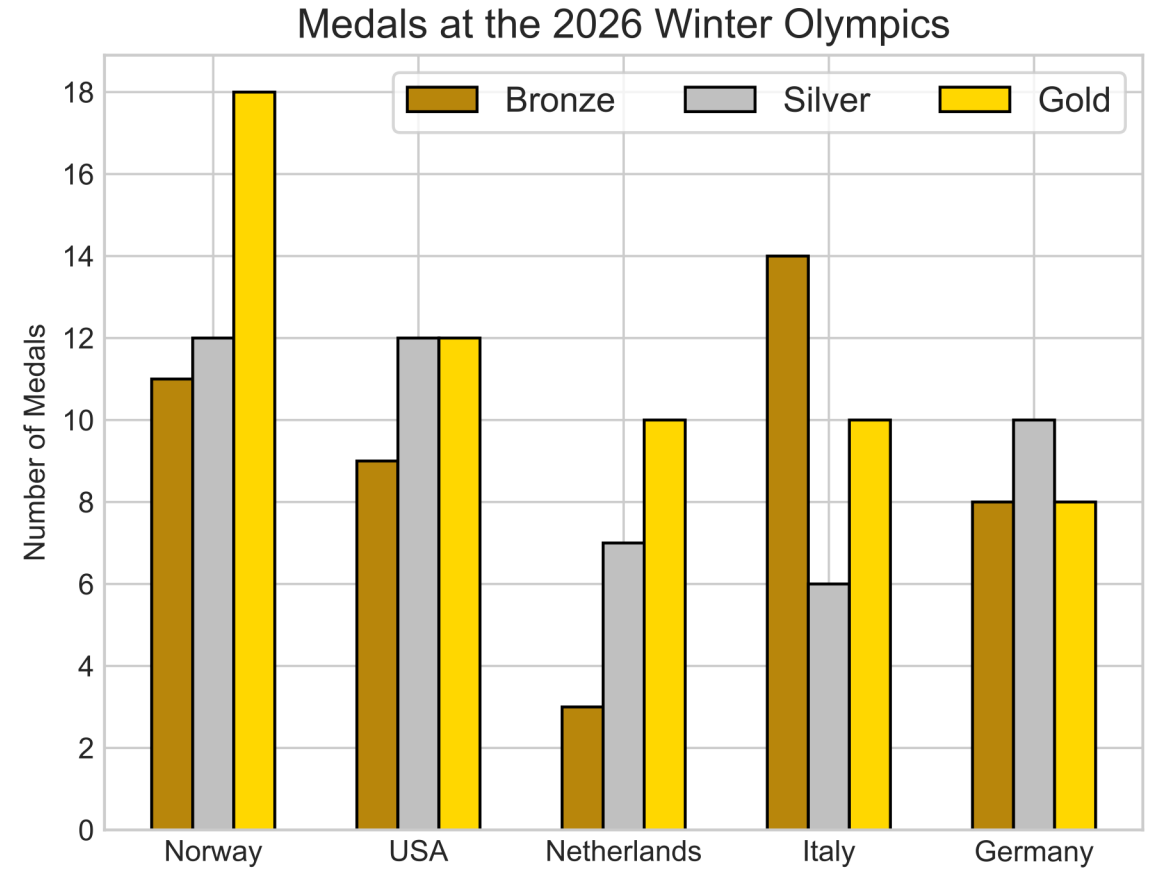
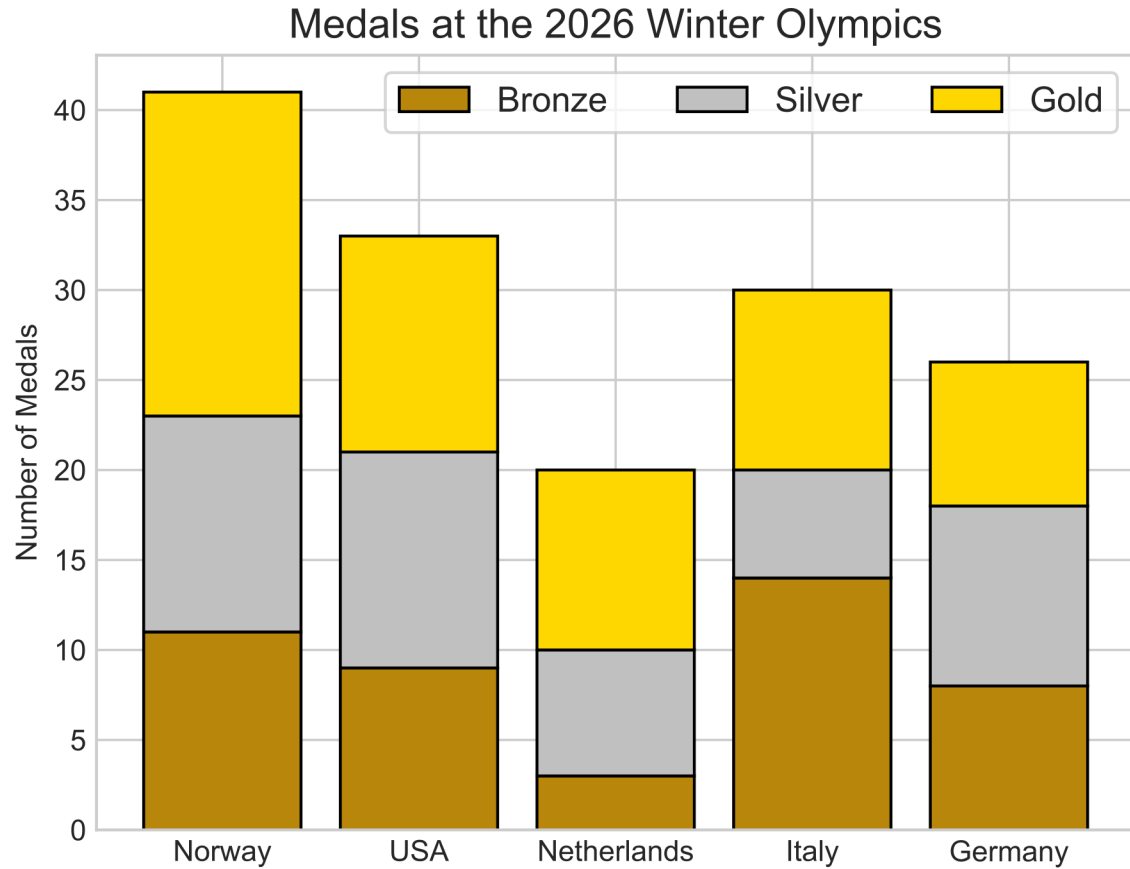
2. Bar Charts



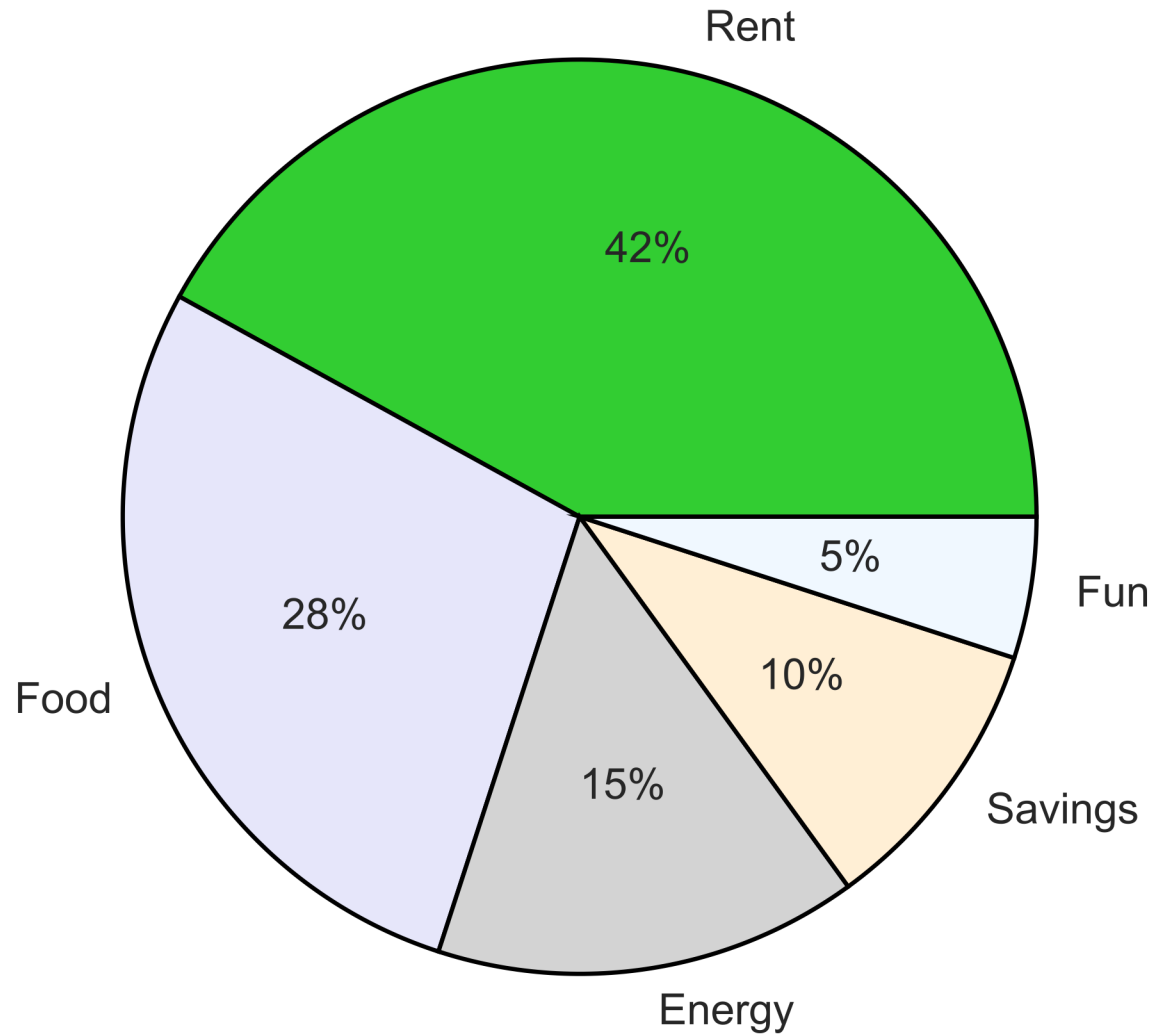
2. Caution! Stacked Bar Charts



2. Caution! Stacked Bar Charts

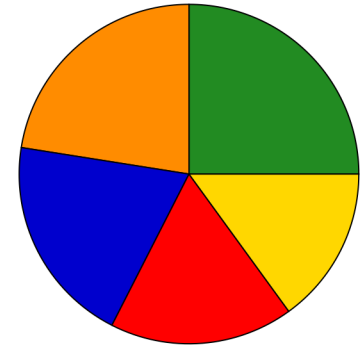
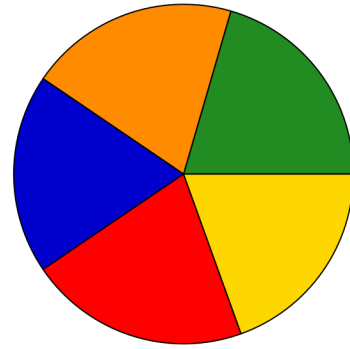
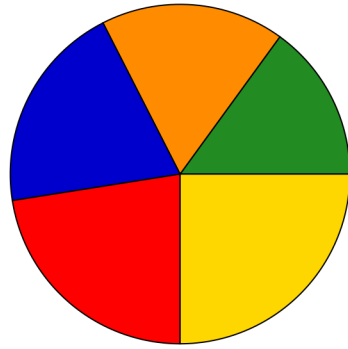


2. Pie Charts

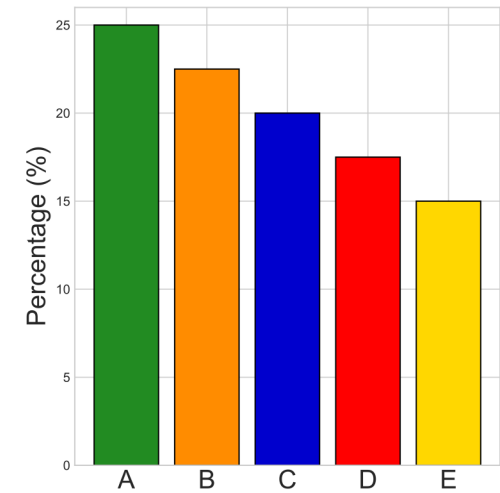
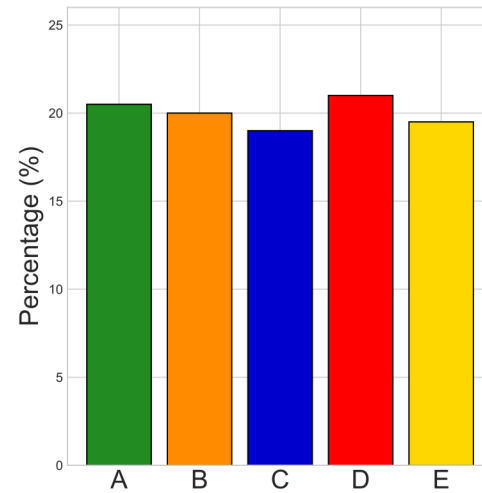
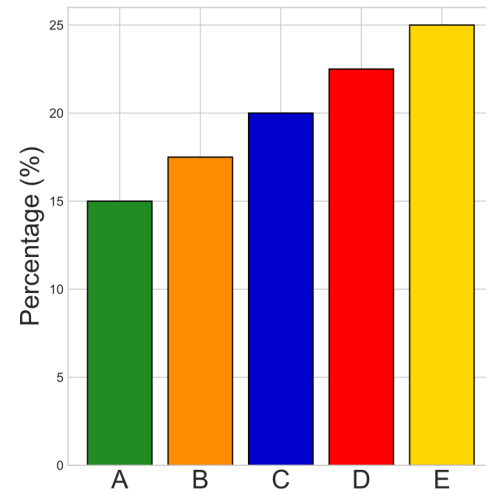
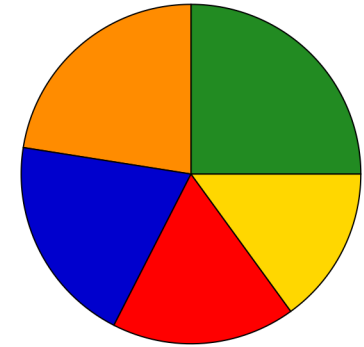
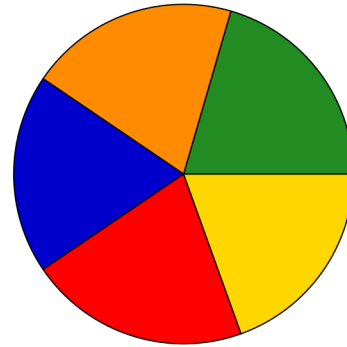
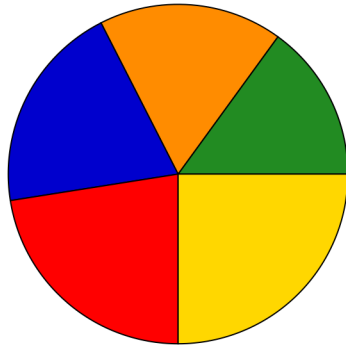


nominal
VS
percentage (ratio)

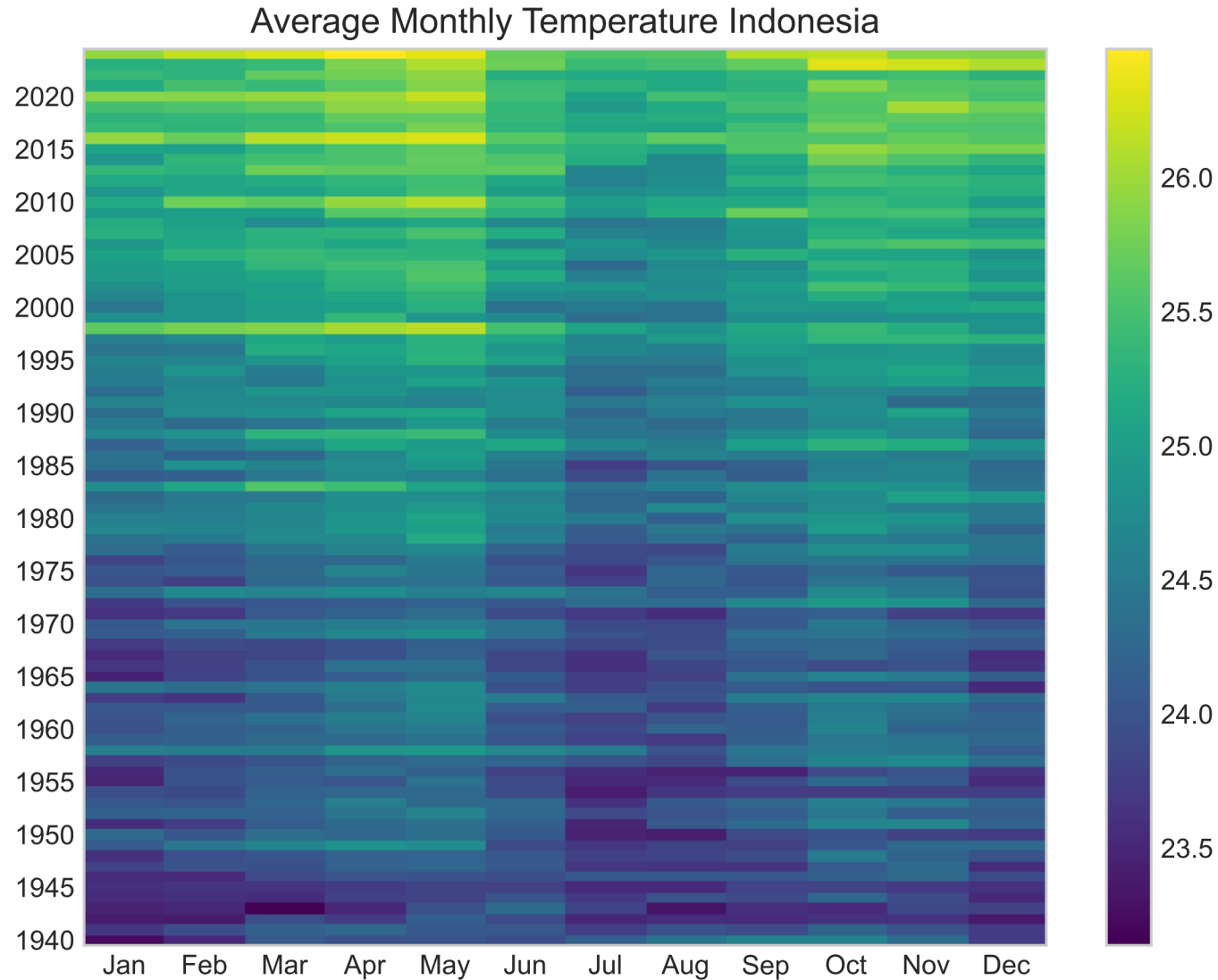
2. Caution! Pie Charts



2. Caution! Pie Charts



2. Heatmaps

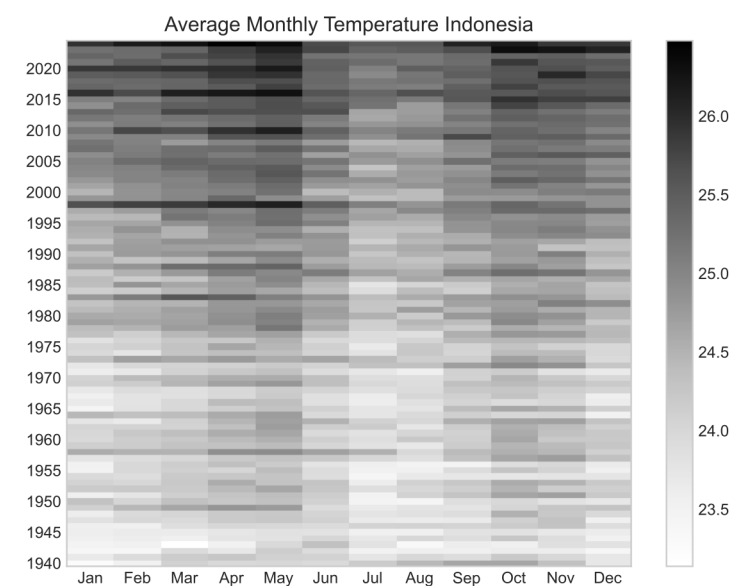
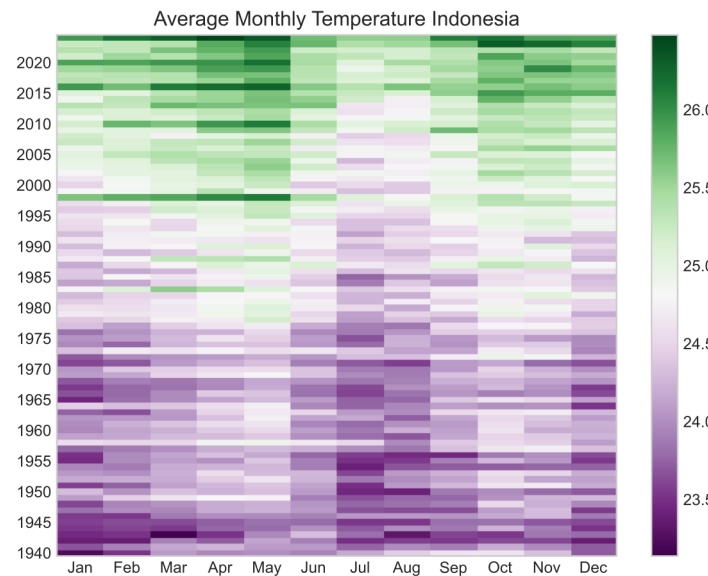
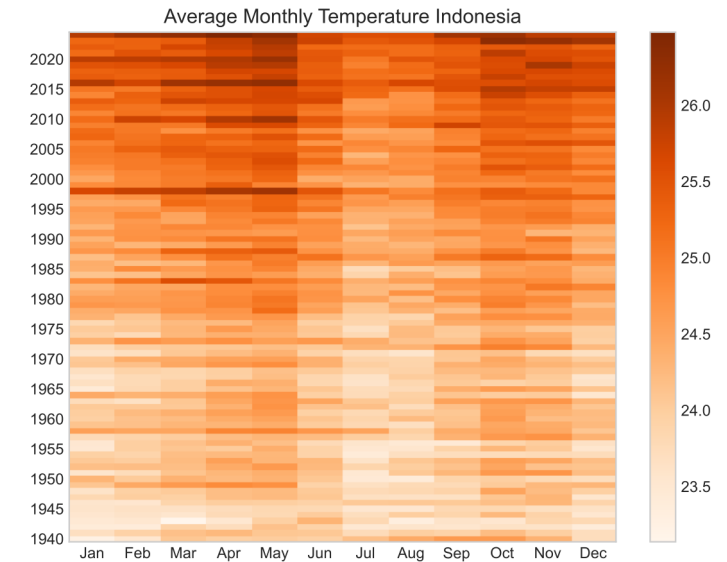
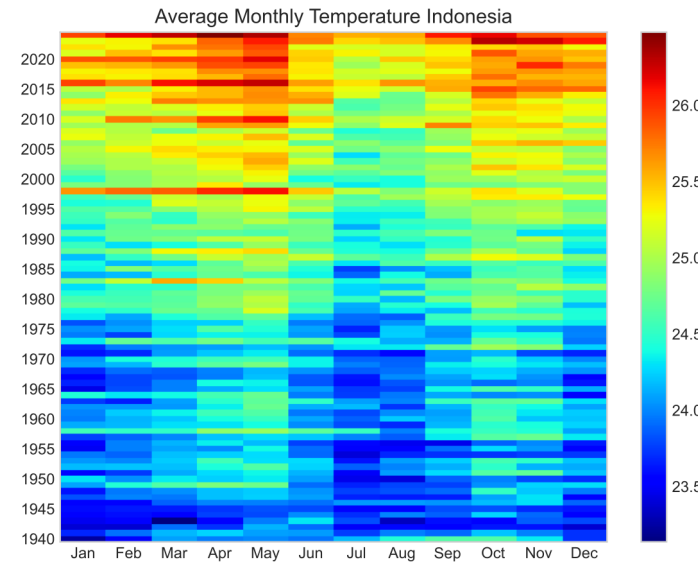


ordinal/interval/ratio
grid
vs
ratio

2. Colourmaps

Colour is difficult to perceive!

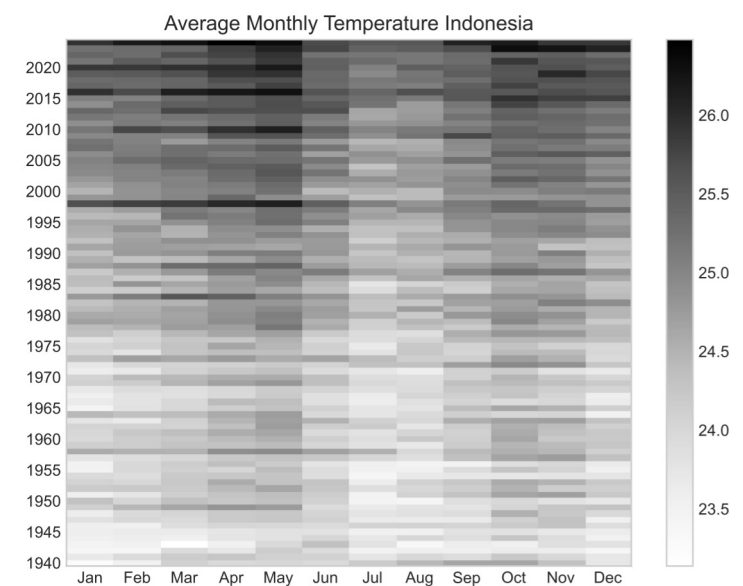
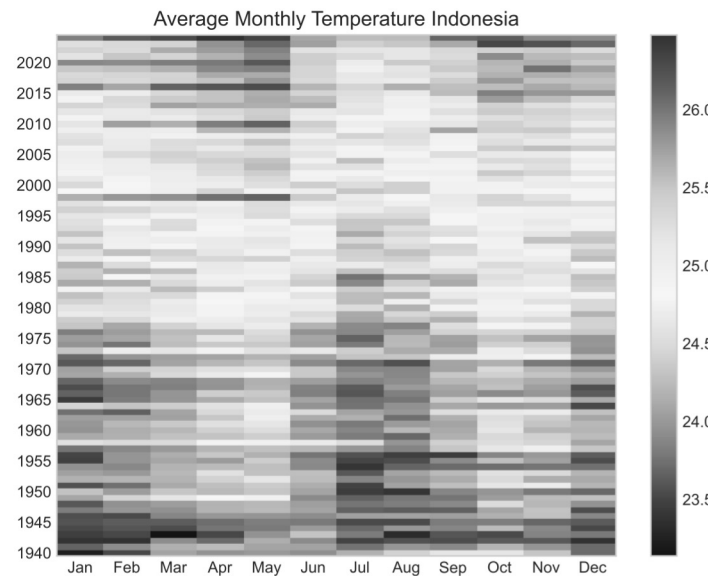
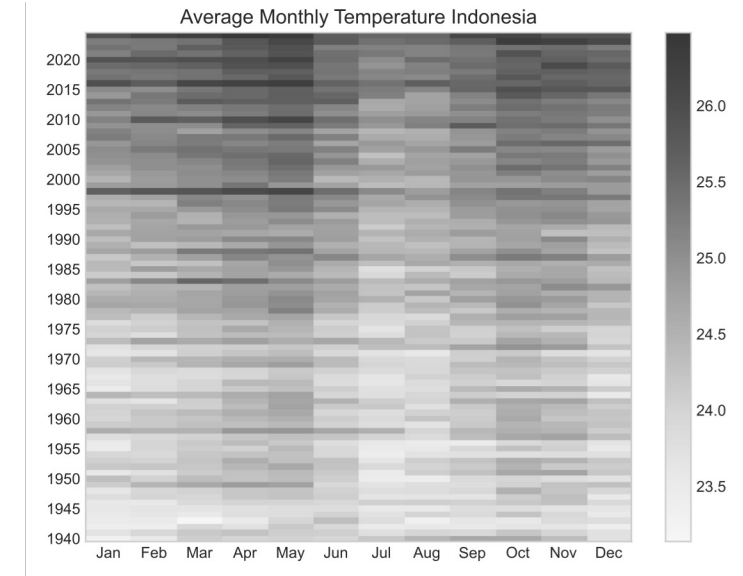
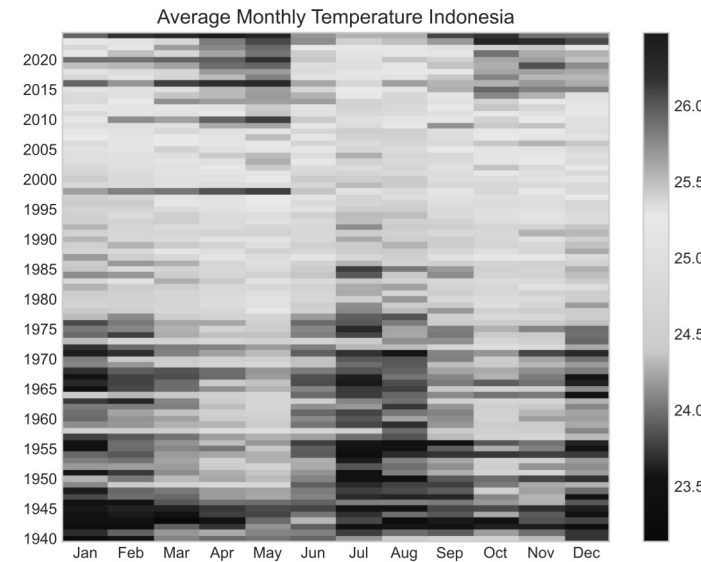
- Individual differences in rods & sensitivity
- Colour blindness
- Screen settings
- Printers & Ink



2. Colourmaps

Colour is difficult to perceive!

- Individual differences in rods & sensitivity
- Colour blindness
- Screen settings
- Printers & Ink



2. Colourmaps

Perceptually Uniform Sequential colormaps

viridis



plasma



inferno



magma

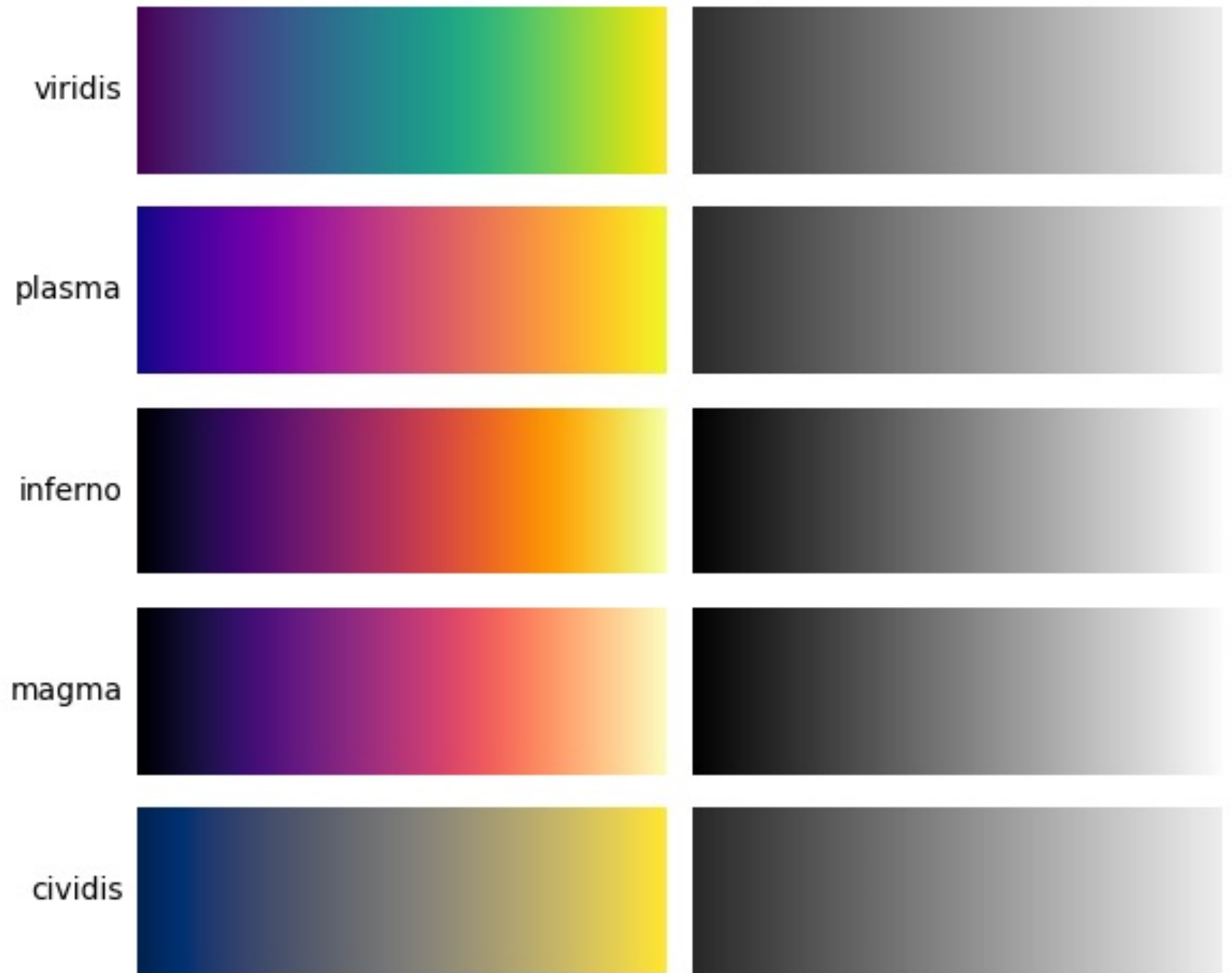


cividis

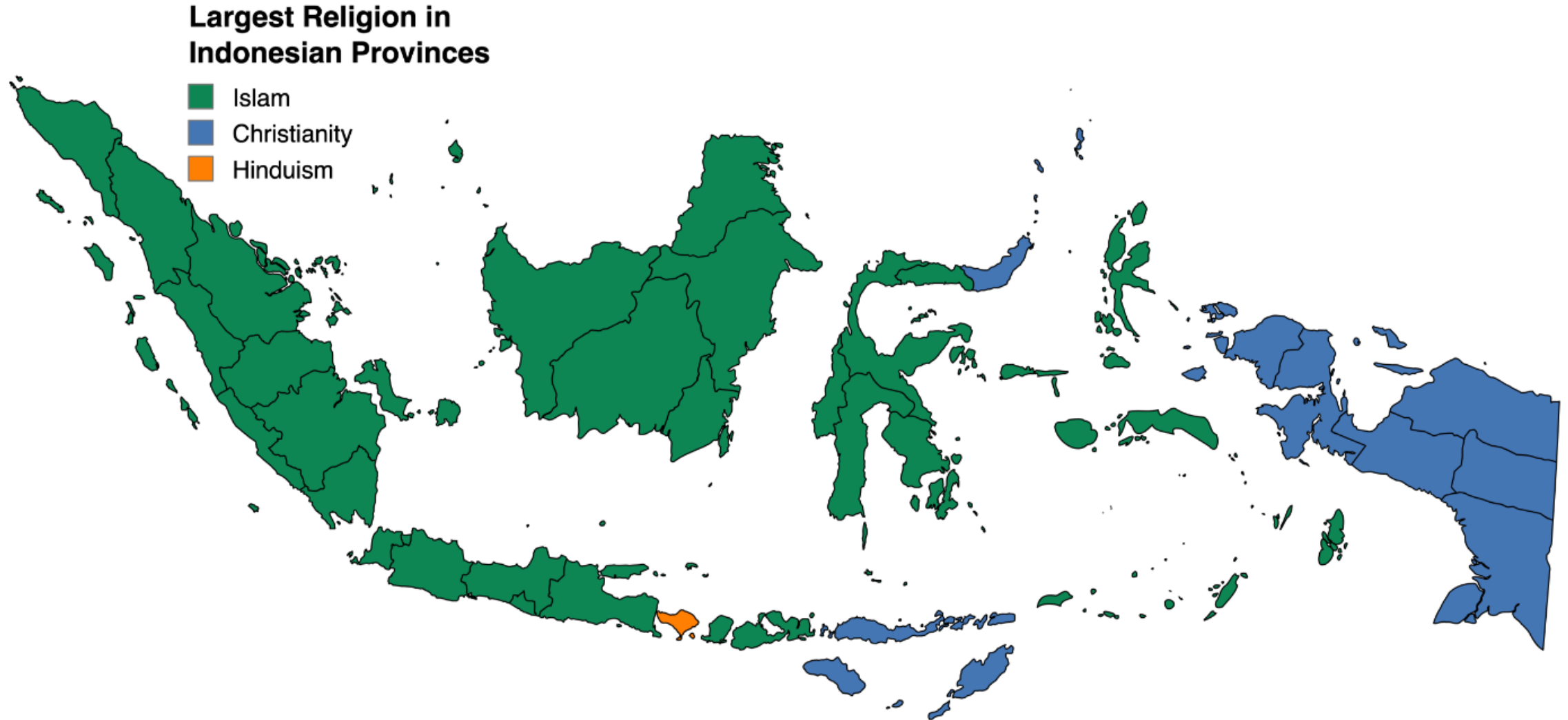


2. Colourmaps

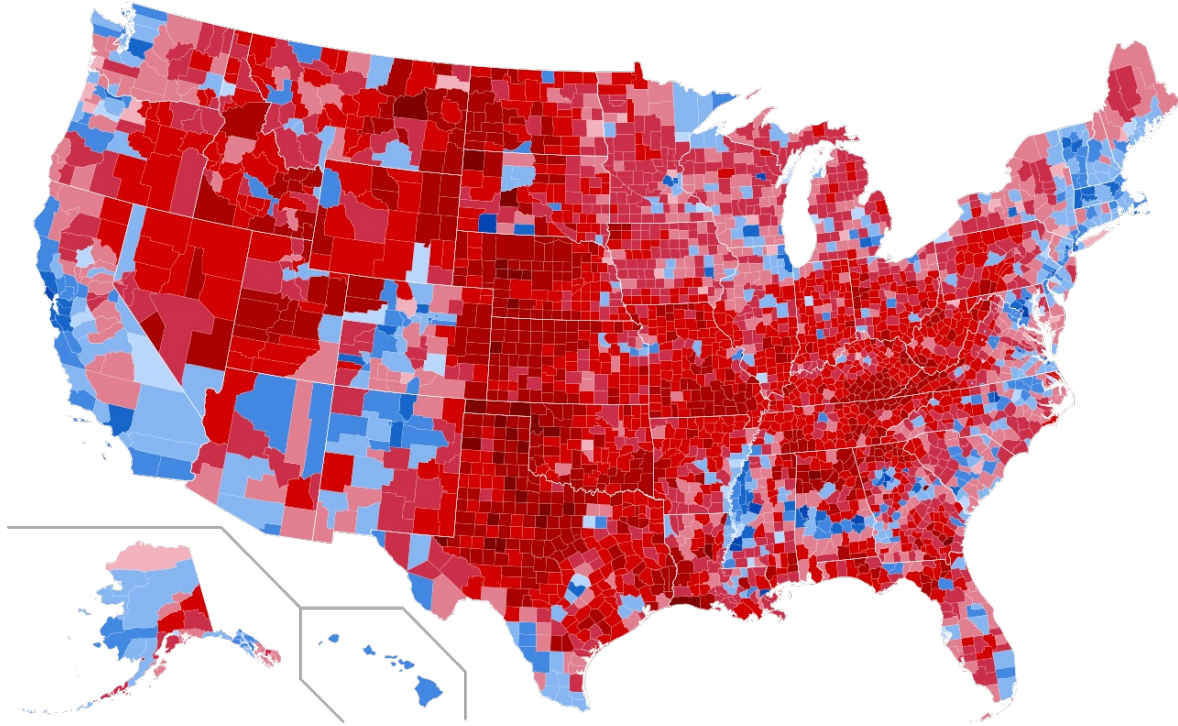
Perceptually Uniform Sequential colormaps



2. Choropleth Maps

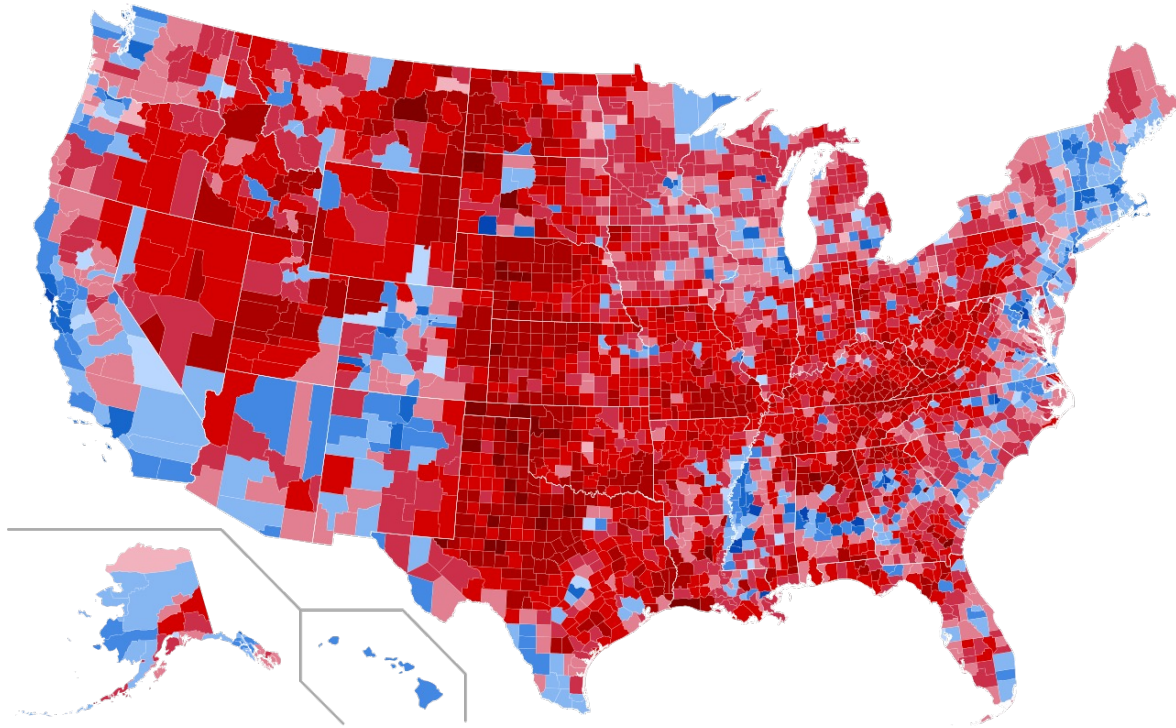


2. Caution! Chloropleth Maps

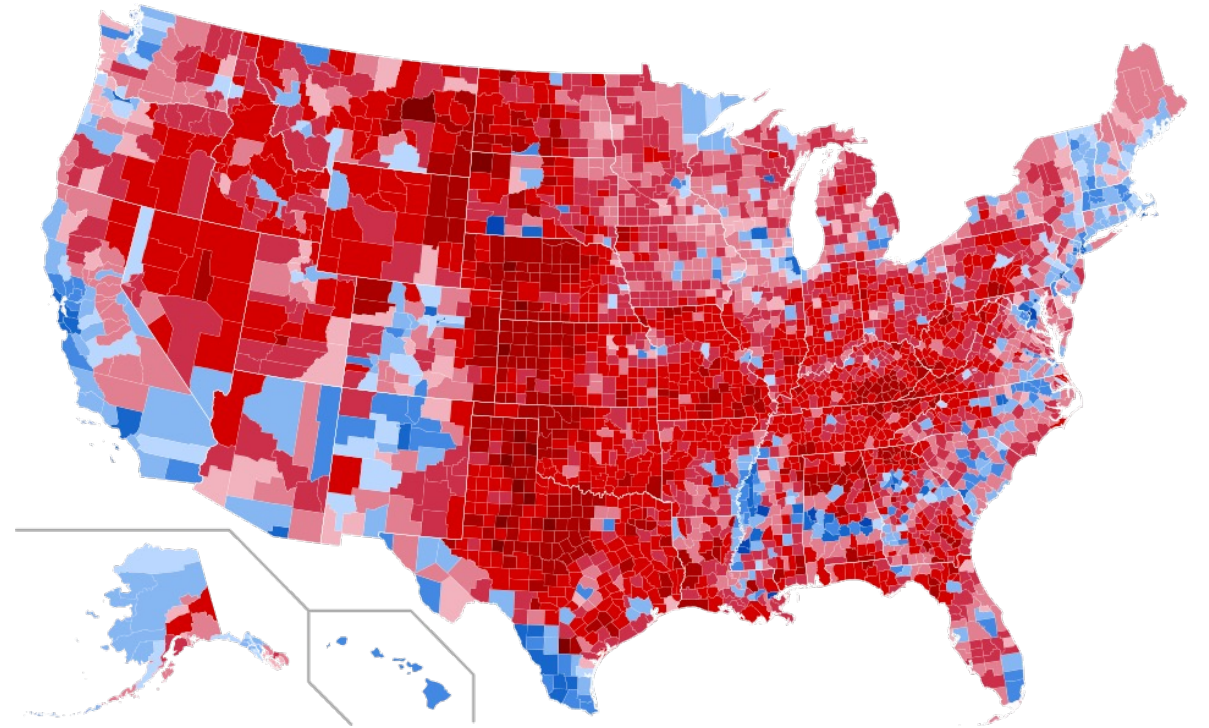


US 2020 Election Map
(Blue won!)

2. Caution! Chloropleth Maps



US 2020 Election Map
(Blue won!)



US 2016 Election Map
(Red won!)

3) Exploring 1 Ratio Variable

3. Histogram

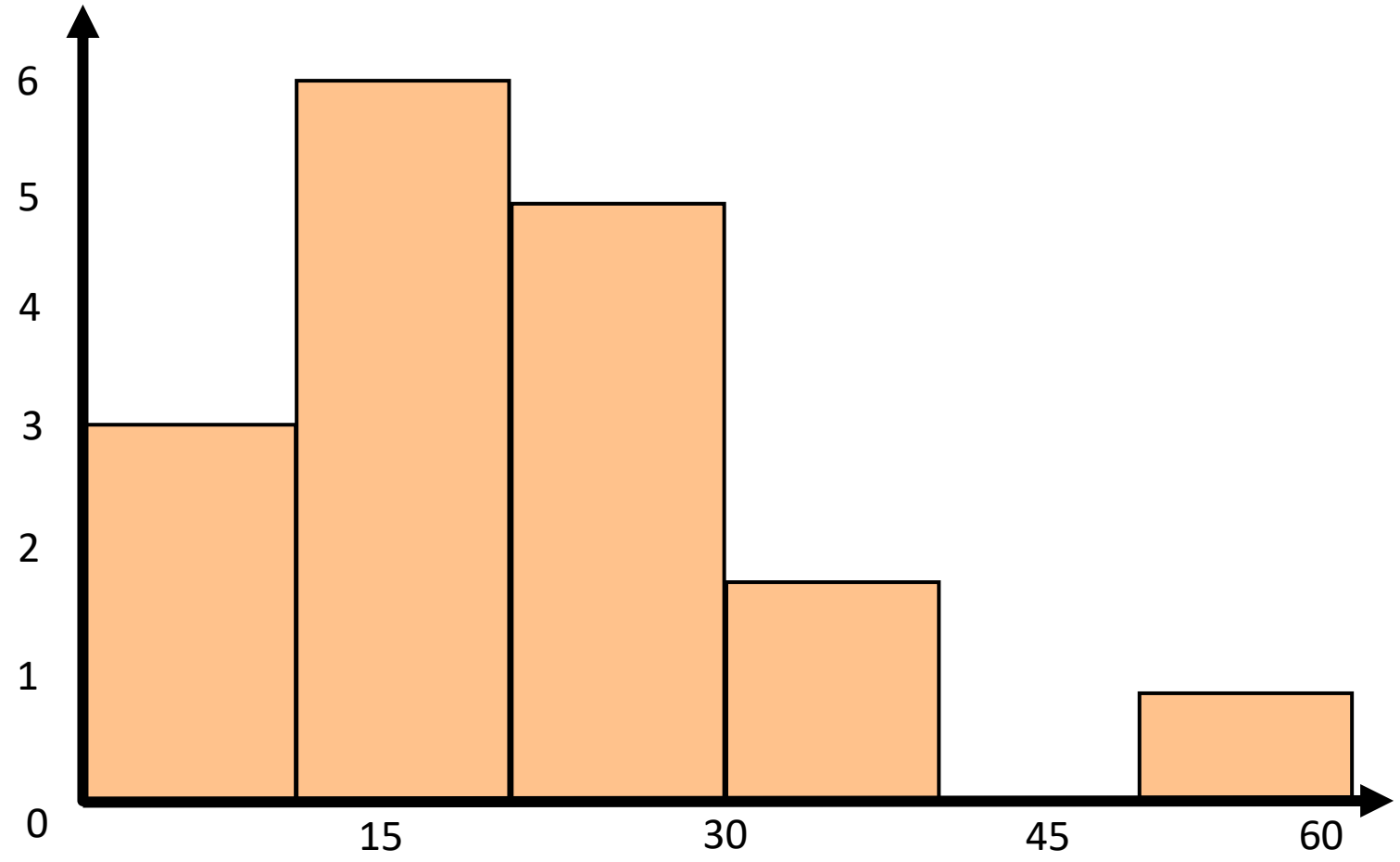
17, 4, 57,
13, 26, 25,
27, 22, 12,
2, 39, 13,
31, 16, 2,
15, 26

3. Histogram

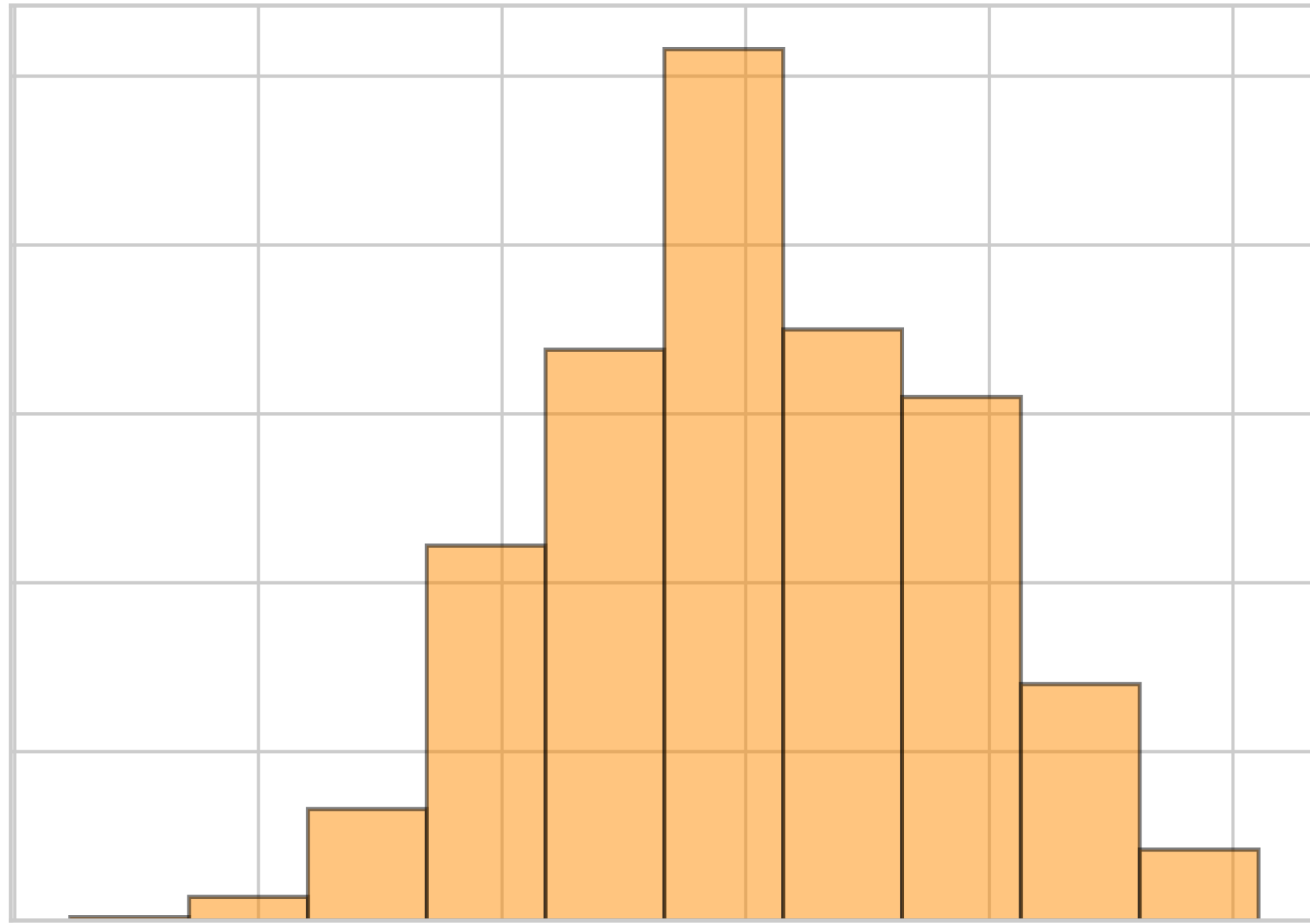
~~17,~~ ~~4,~~ ~~57,~~
~~13,~~ ~~26,~~ ~~25,~~
~~27,~~ ~~22,~~ ~~12,~~
~~2,~~ ~~39,~~ ~~13,~~
~~31,~~ ~~16,~~ ~~2,~~
~~15,~~ ~~26~~

Bin	Tally
0 – 10	
10 – 20	
20 – 30	
30 – 40	
40 – 50	
50 - 60	

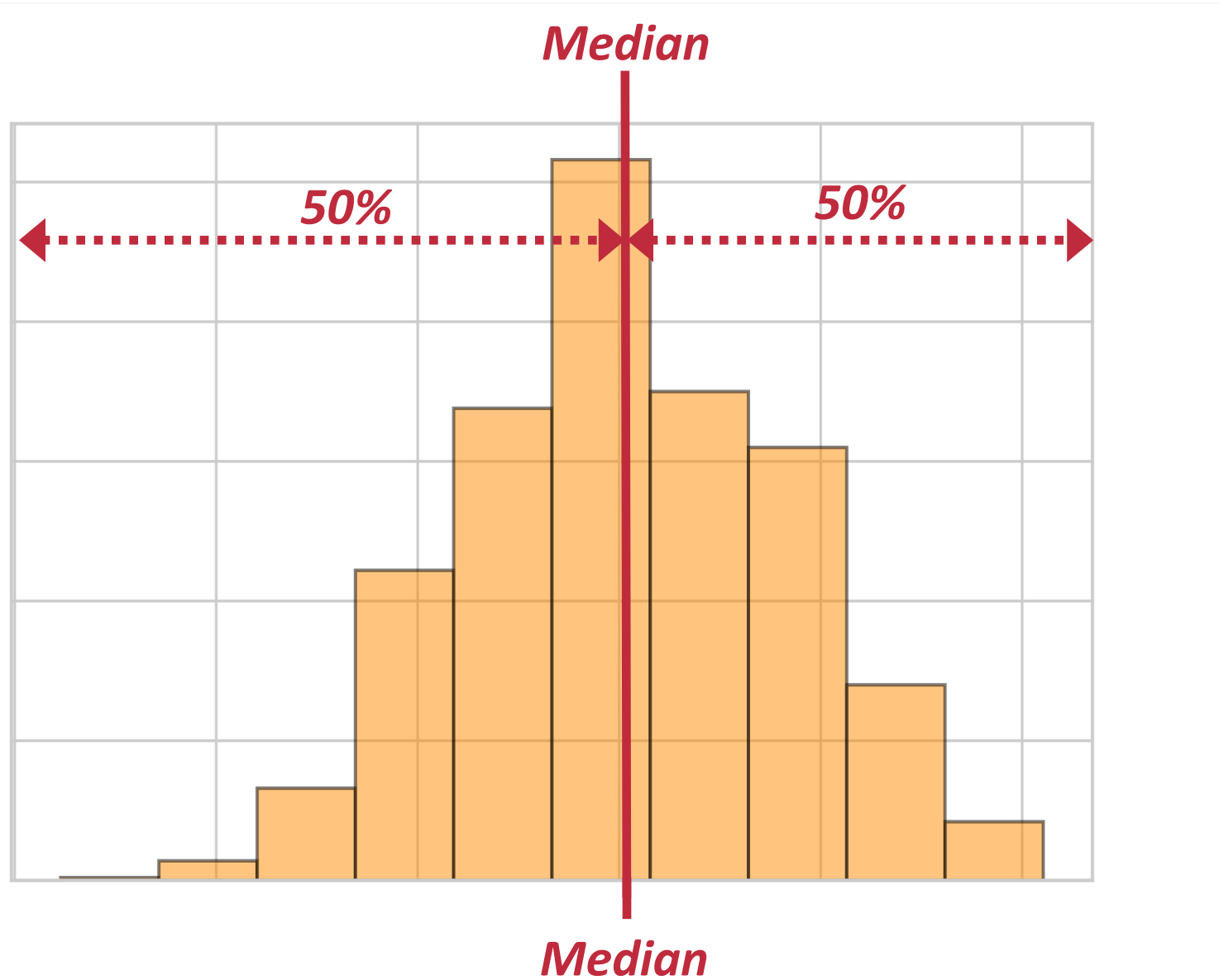
3. Histogram



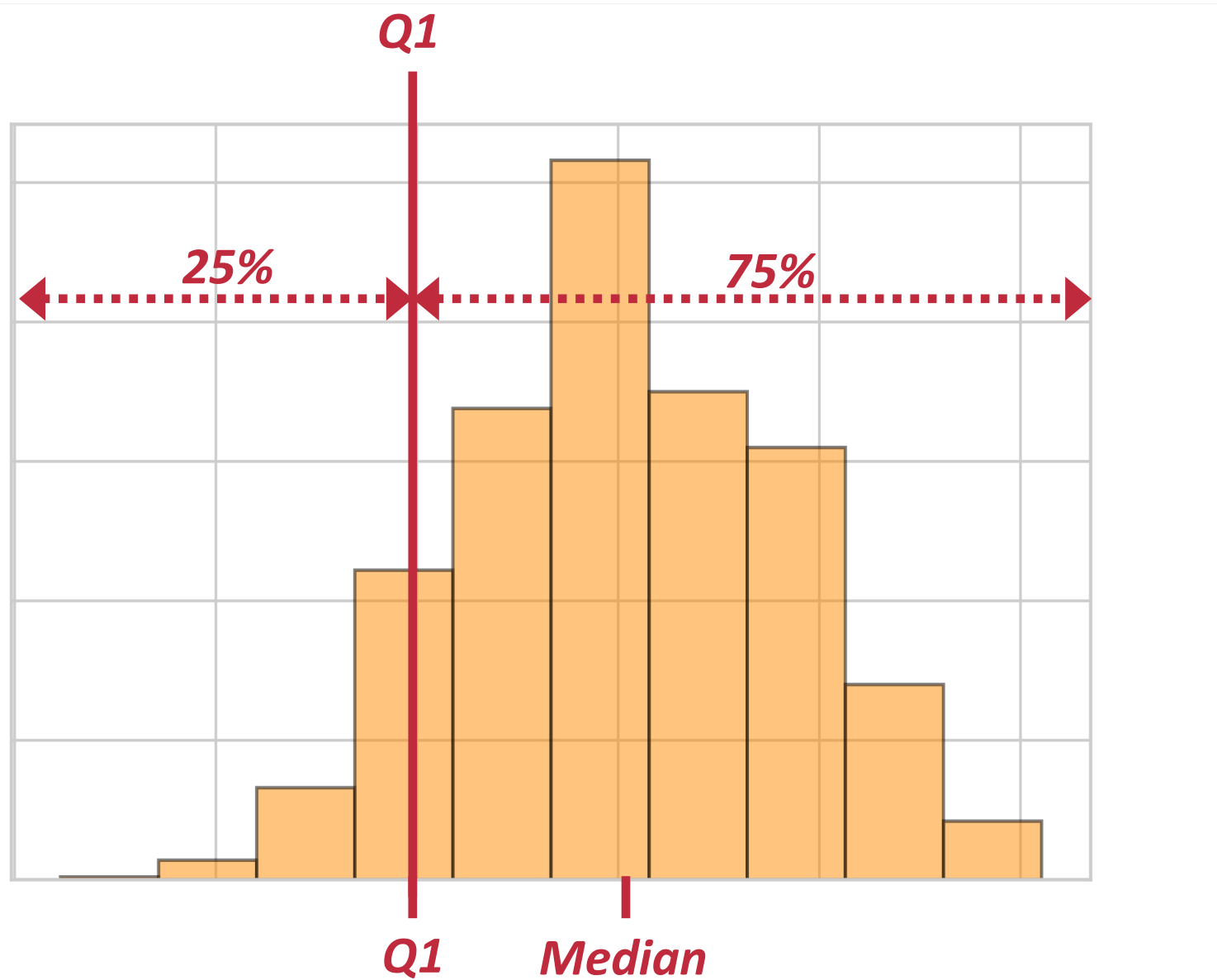
3. Quartiles



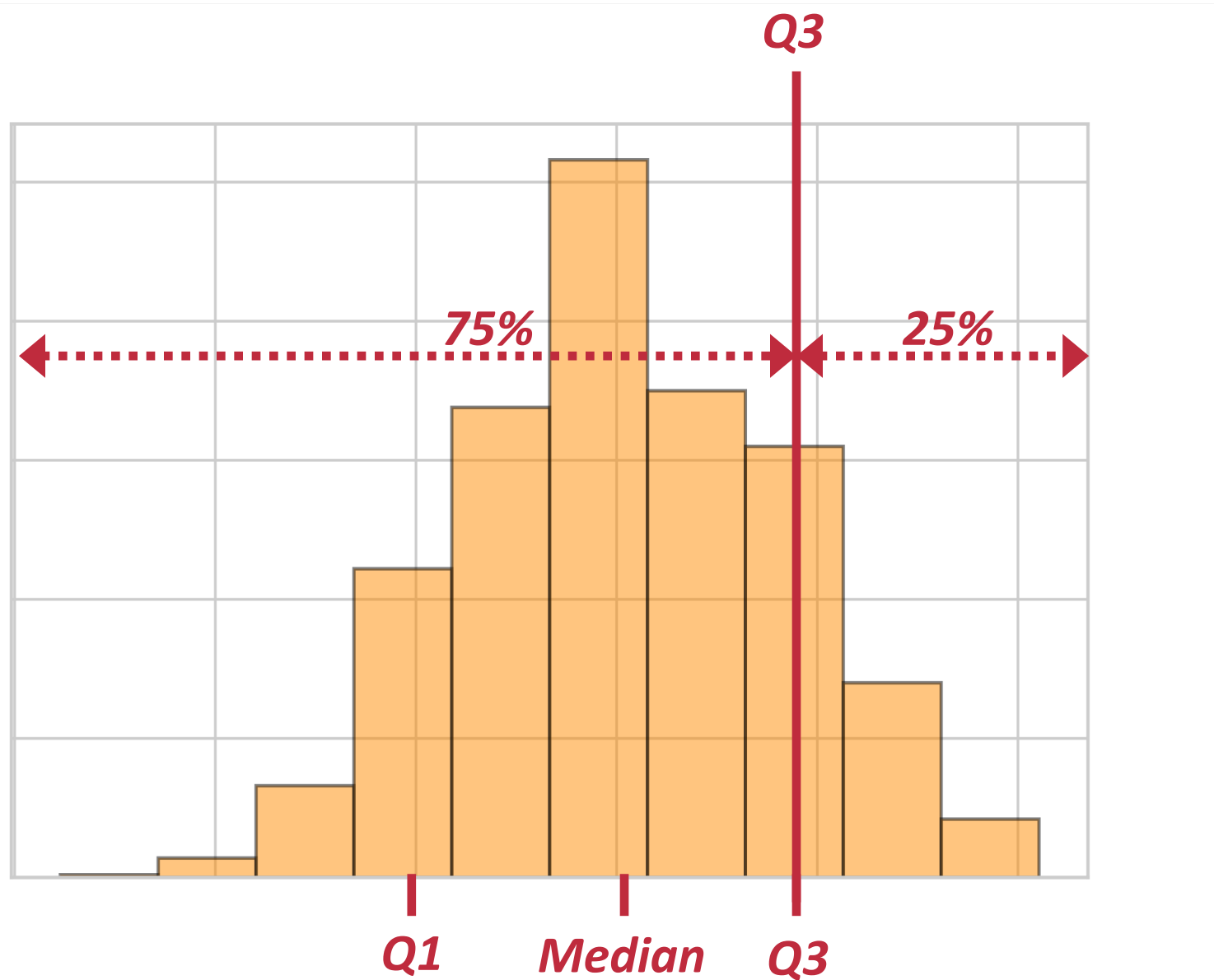
3. Quartiles



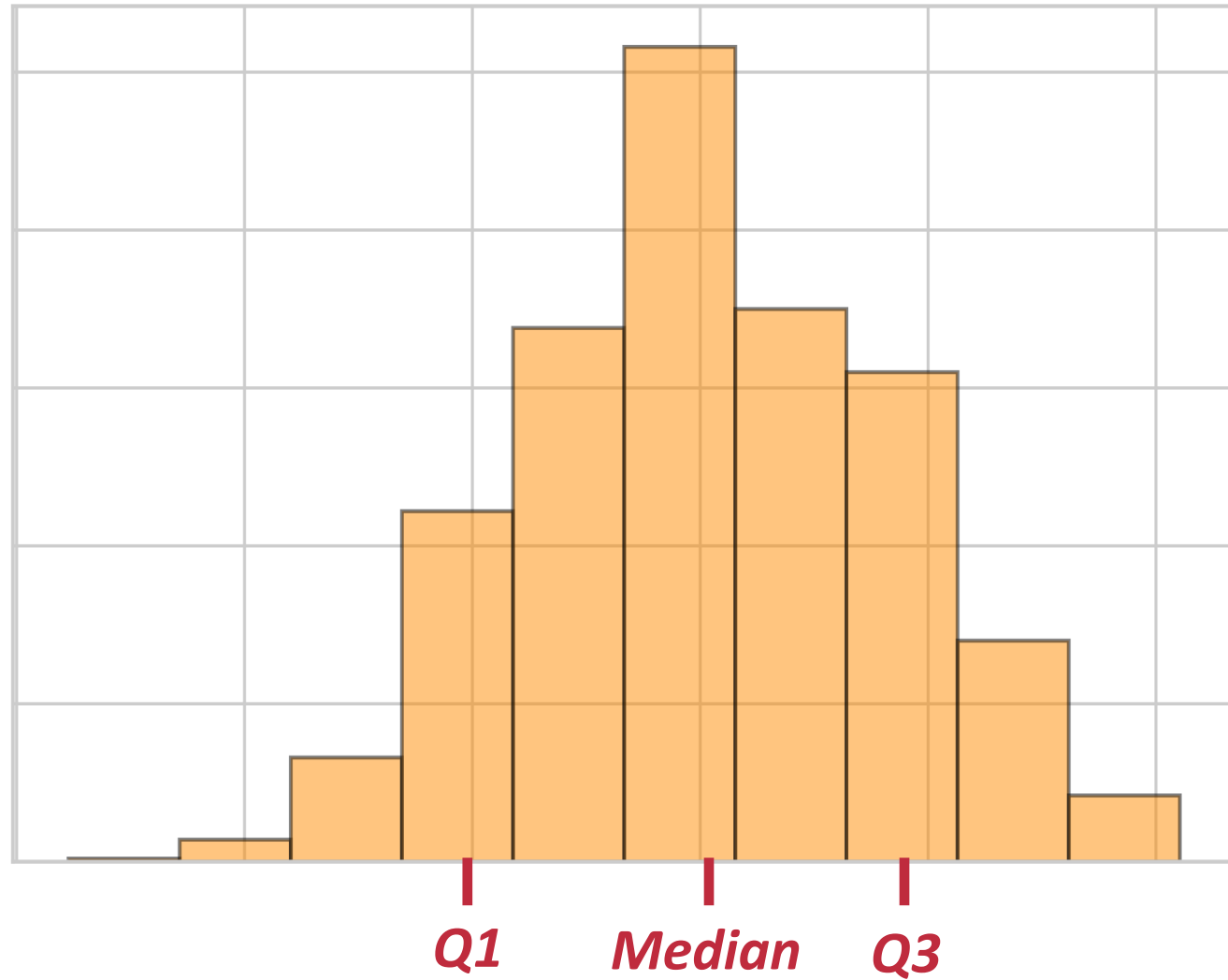
3. Quartiles



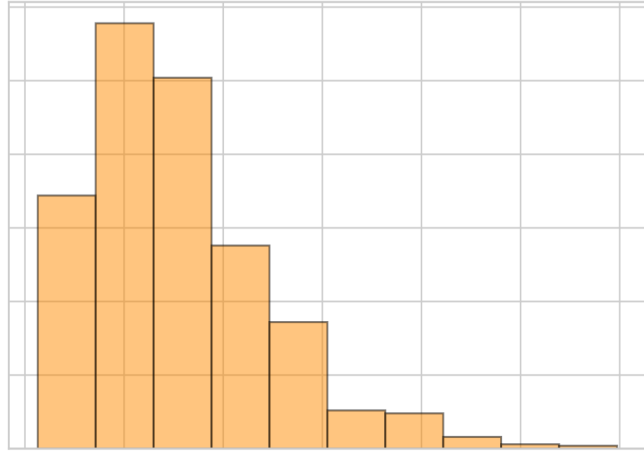
3. Quartiles



3. Quartiles

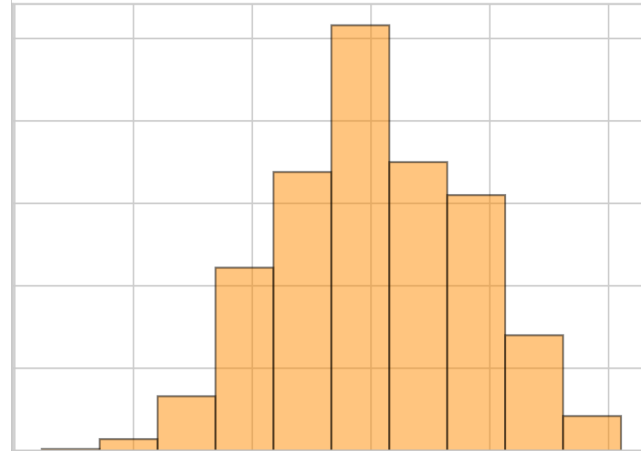


3. Skewness



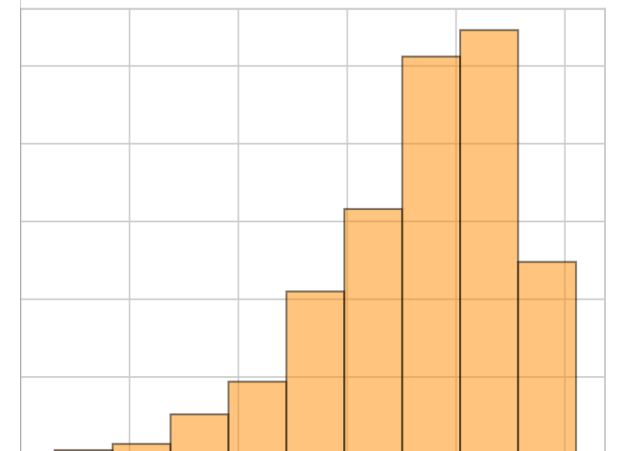
Positive skew

median < mean



No skew

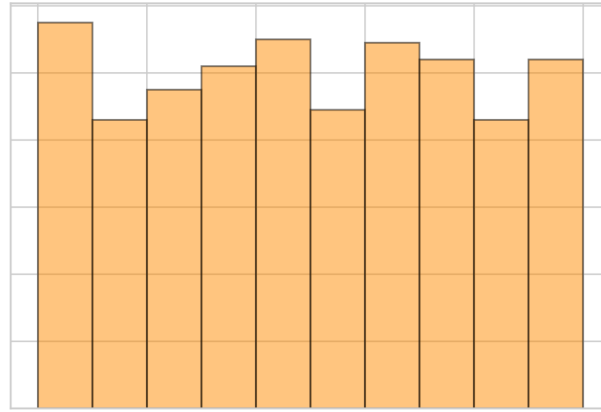
median = mean



Negative skew

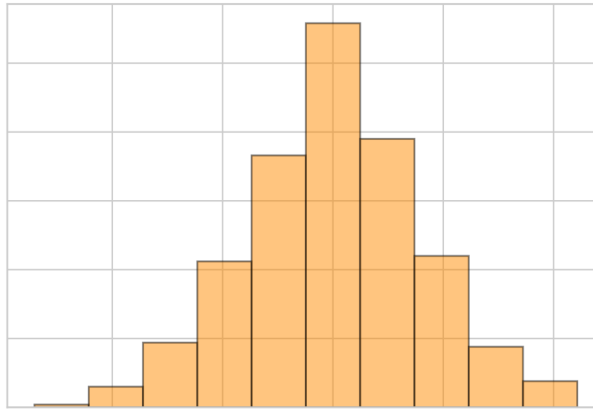
median > mean

3. Kurtosis

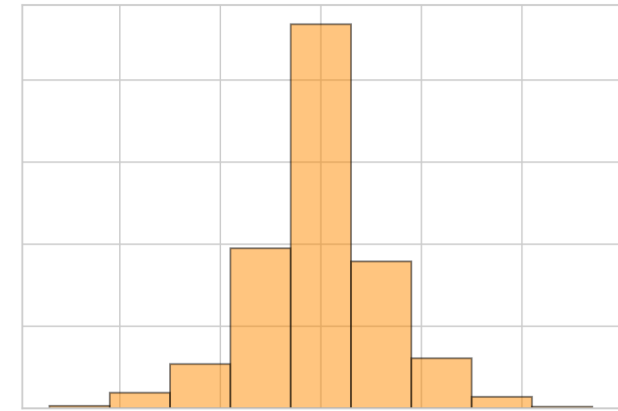


Low kurtosis

Data
concentrated in
the tails



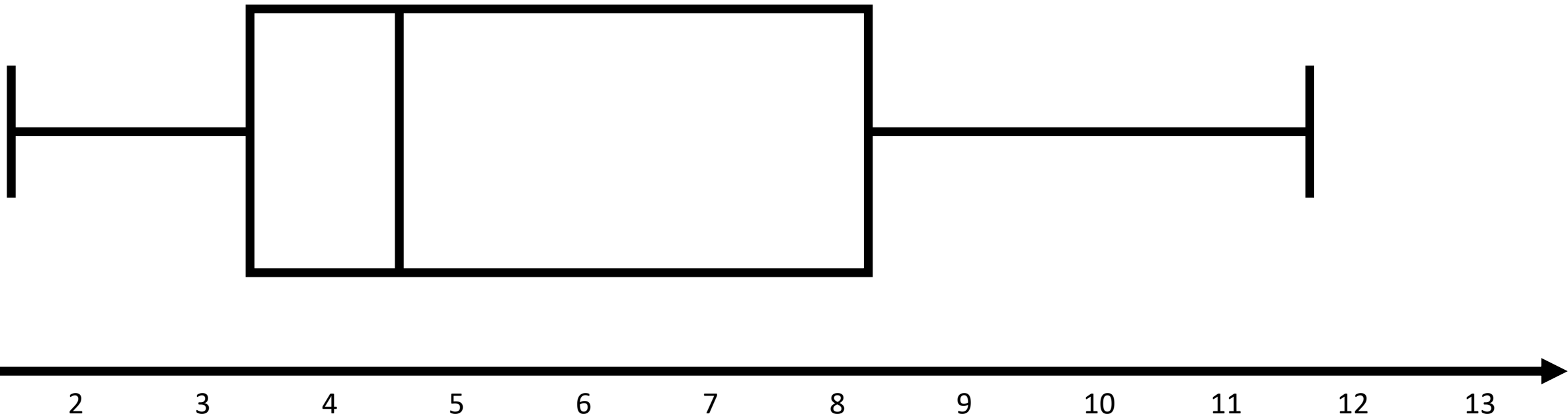
Moderate kurtosis



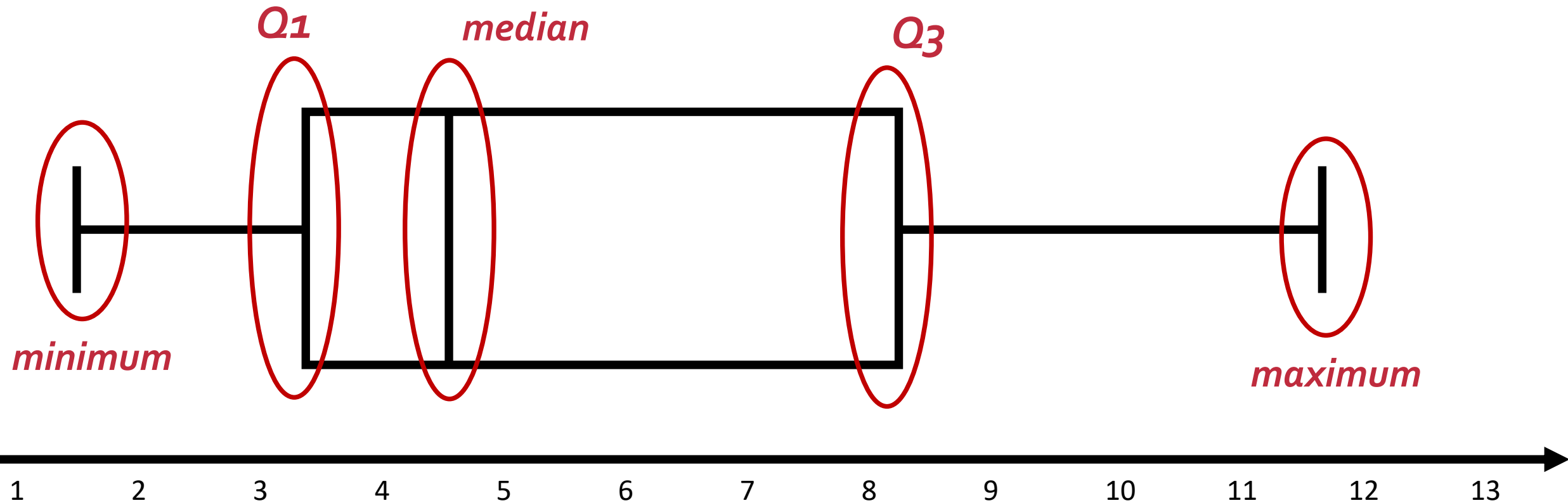
High kurtosis

Data
concentrated
around the mean

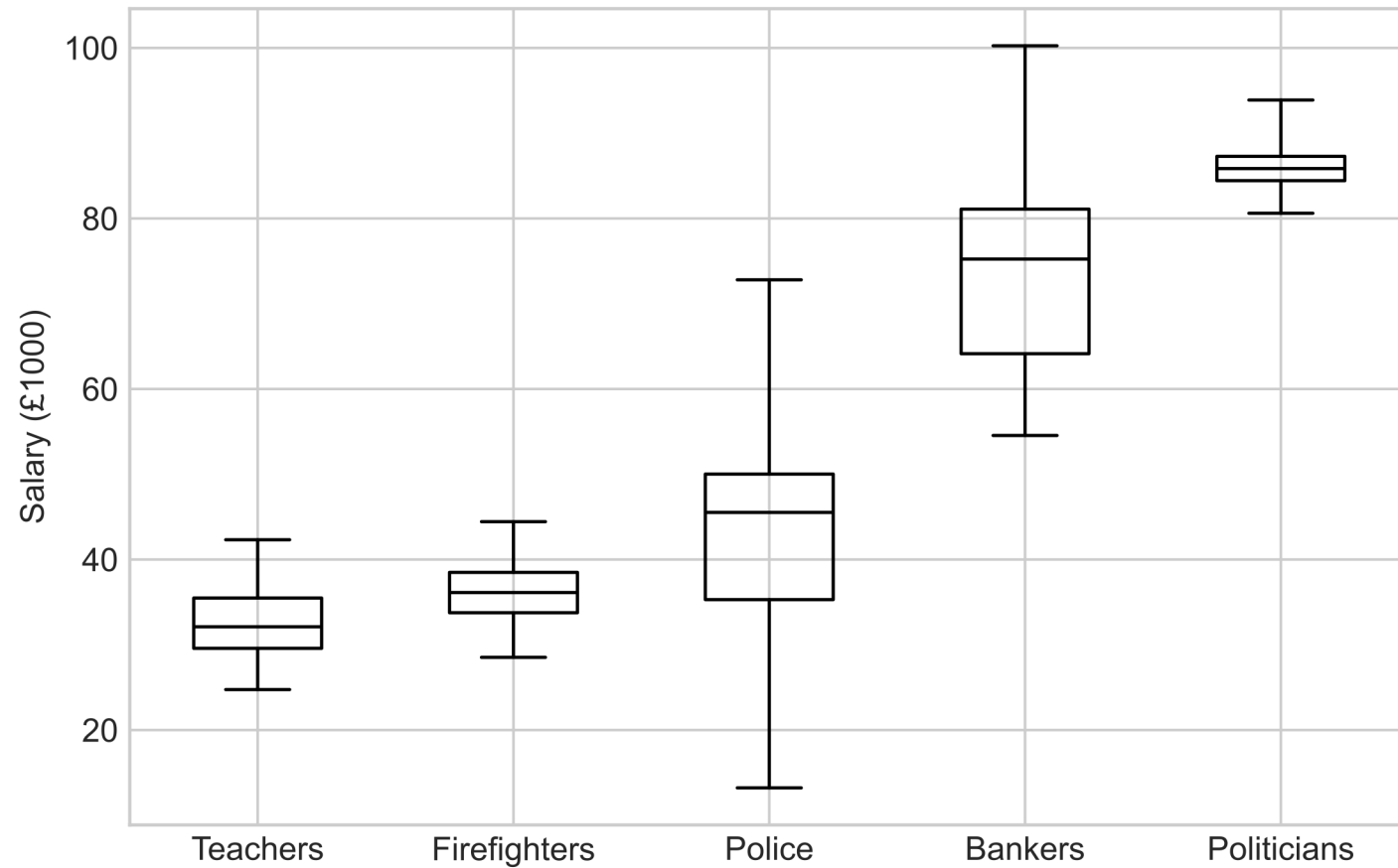
3. Boxplot



3. Boxplot

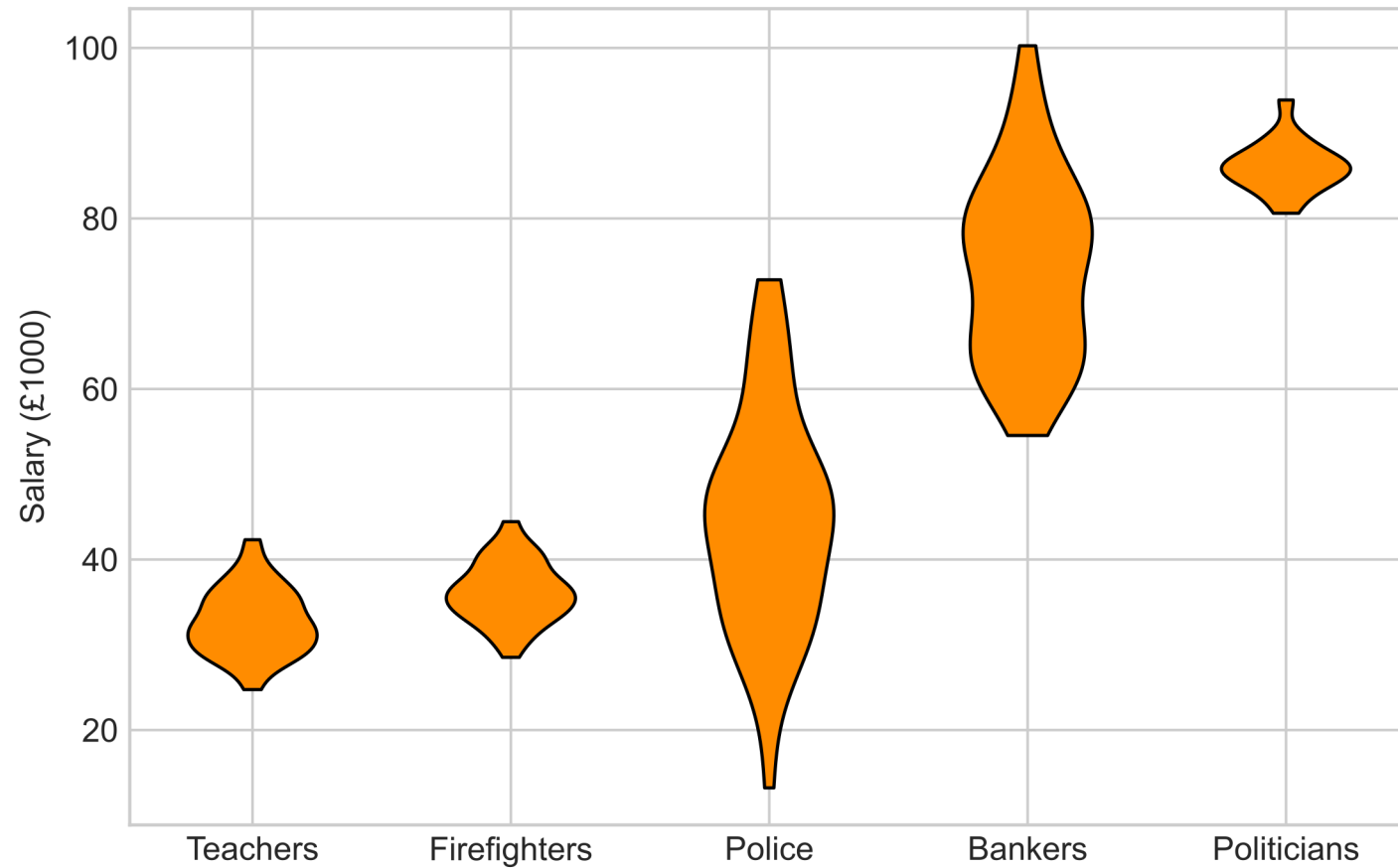


3. Boxplots



nominal/ordinal
VS
sets of ratio

3. Violinplots



nominal/ordinal
VS
sets of ratio

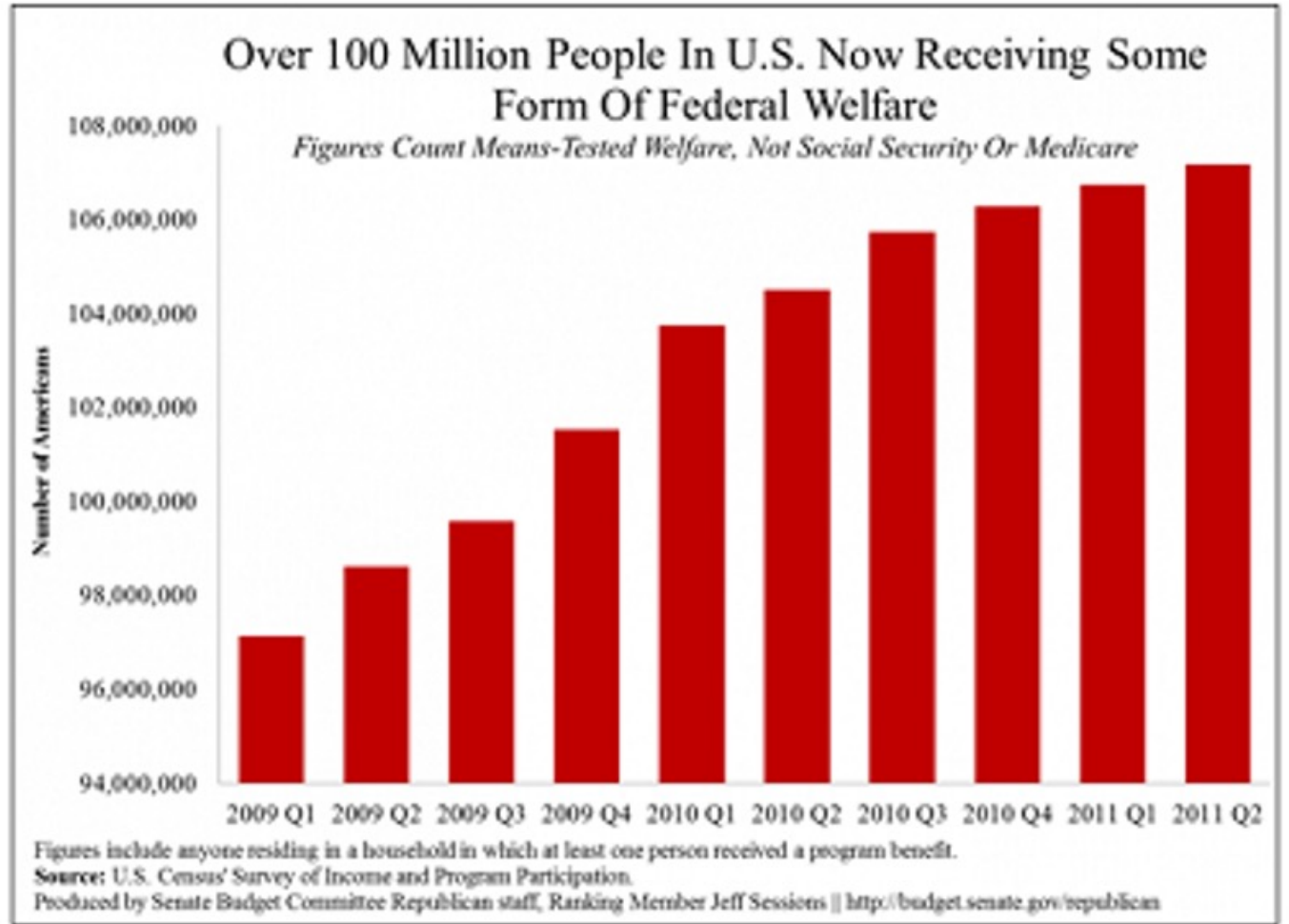
4) Common Pitfalls

4. Common Pitfalls

- **Omitting baseline**
- Misleading scale
- Omitting data
- Breaking conventions
- Distorting reality
- Too many stories
- Too many dimensions
- Too many elements

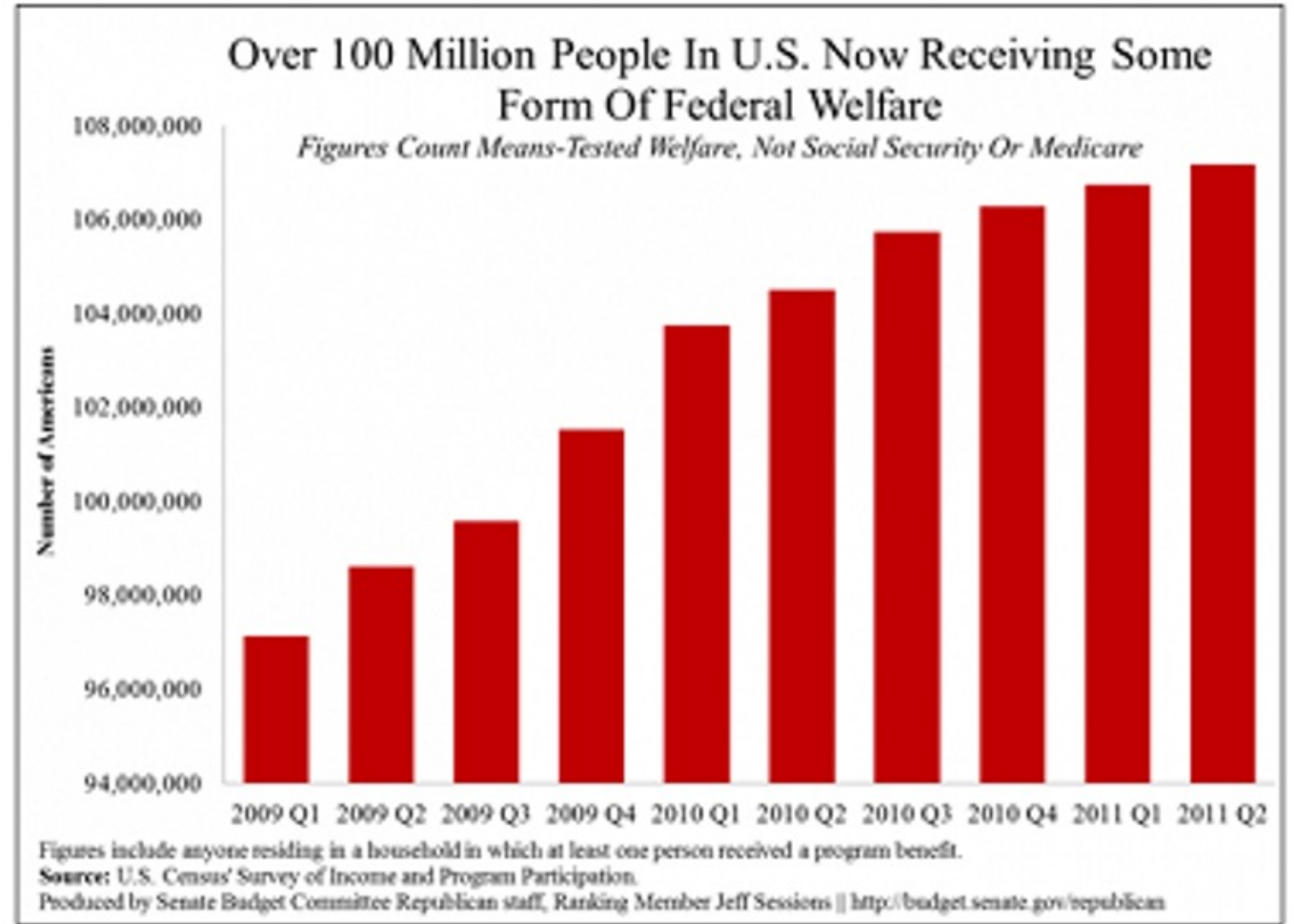
4. Common Pitfalls

- Omitting baseline
- Misleading scale
- Omitting data
- Breaking conventions
- Distorting reality
- Too many stories
- Too many dimensions
- Too many elements



4. Common Pitfalls

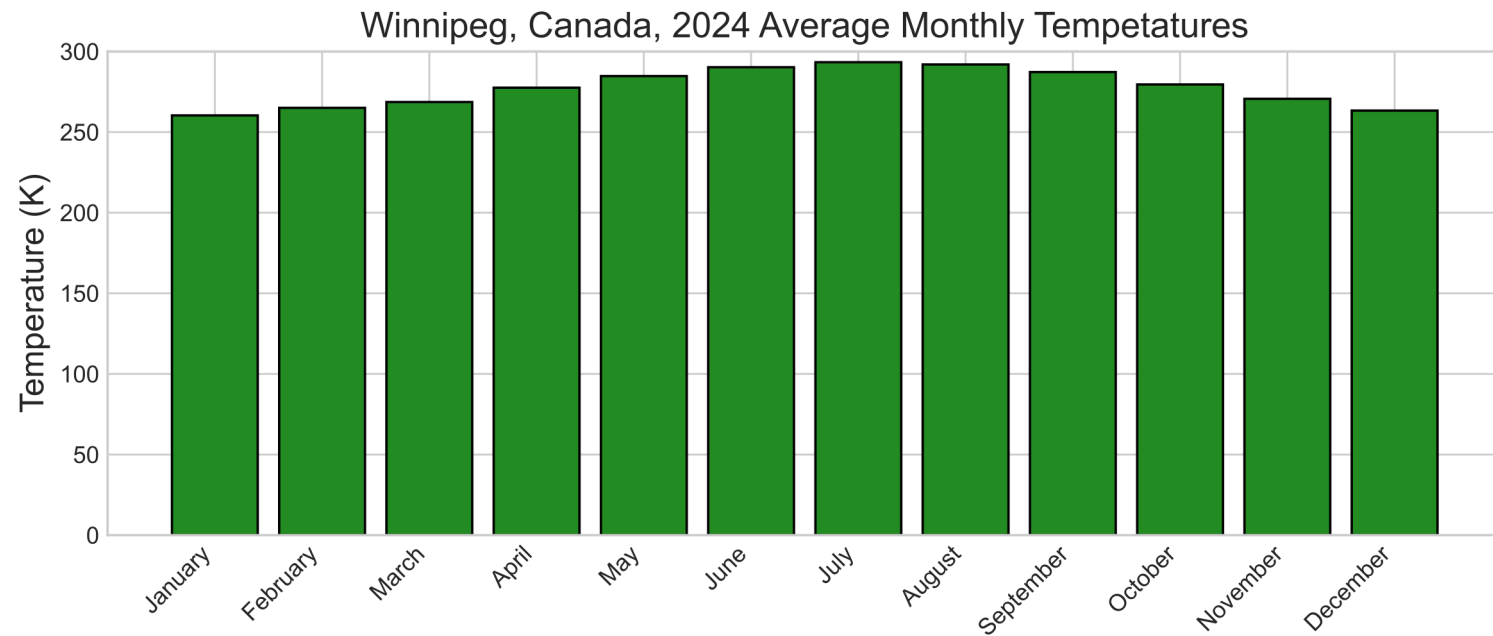
- Omitting baseline
- Misleading scale
- Omitting data
- Breaking conventions
- Distorting reality
- Too many stories
- Too many dimensions
- Too many elements



For ratio data, always show the 0!

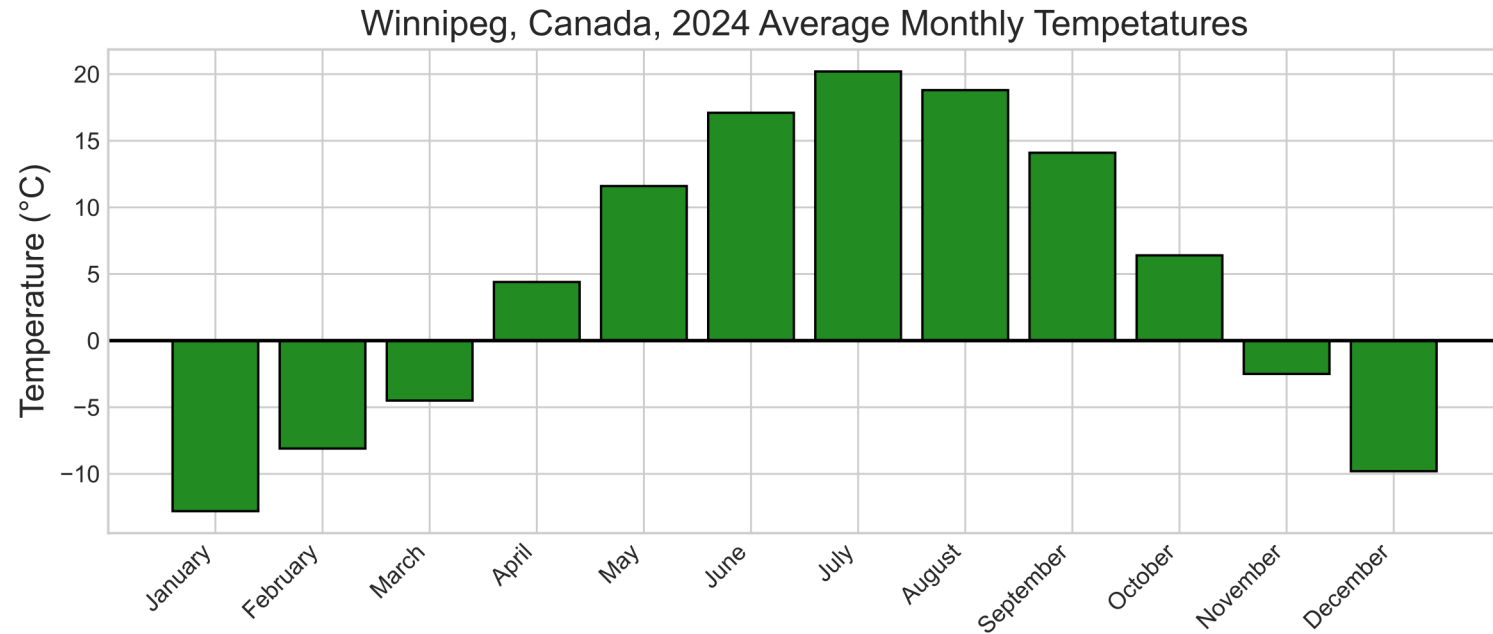
4. Common Pitfalls

- Omitting baseline
- **Misleading scale**
- Omitting data
- Breaking conventions
- Distorting reality
- Too many stories
- Too many dimensions
- Too many elements



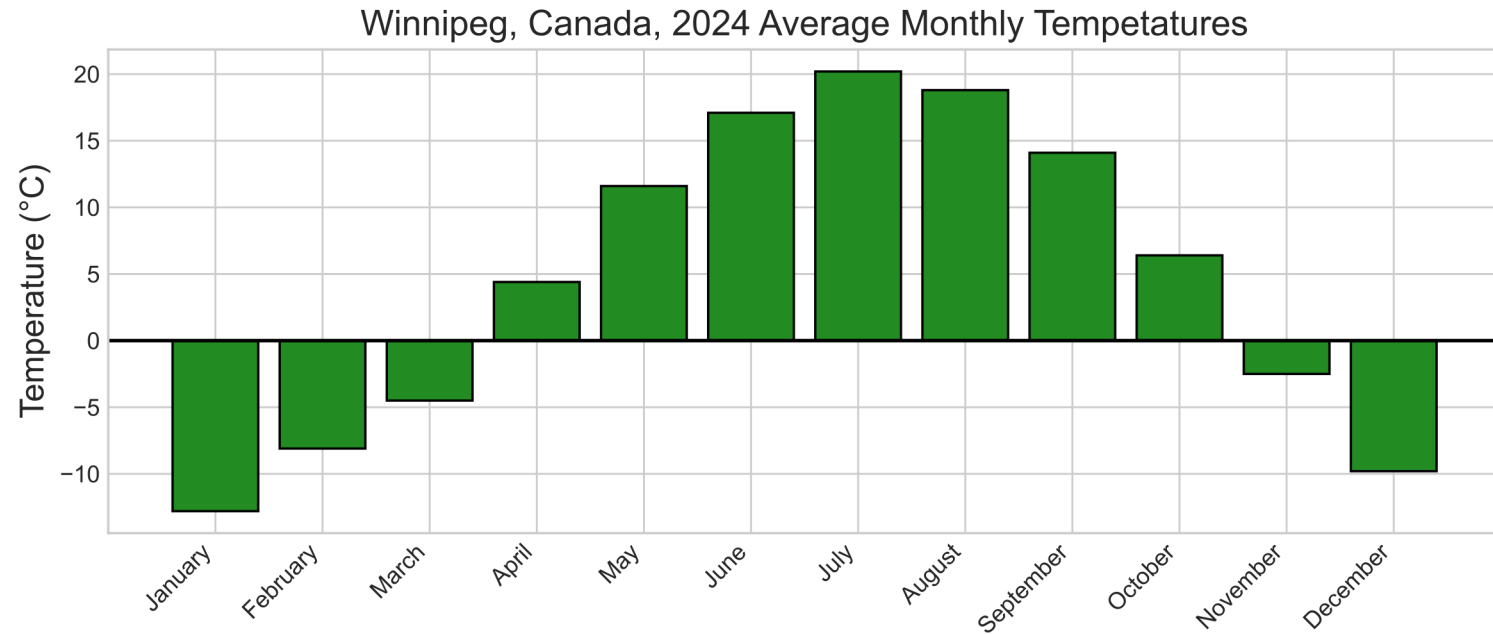
4. Common Pitfalls

- Omitting baseline
- **Misleading scale**
- Omitting data
- Breaking conventions
- Distorting reality
- Too many stories
- Too many dimensions
- Too many elements



4. Common Pitfalls

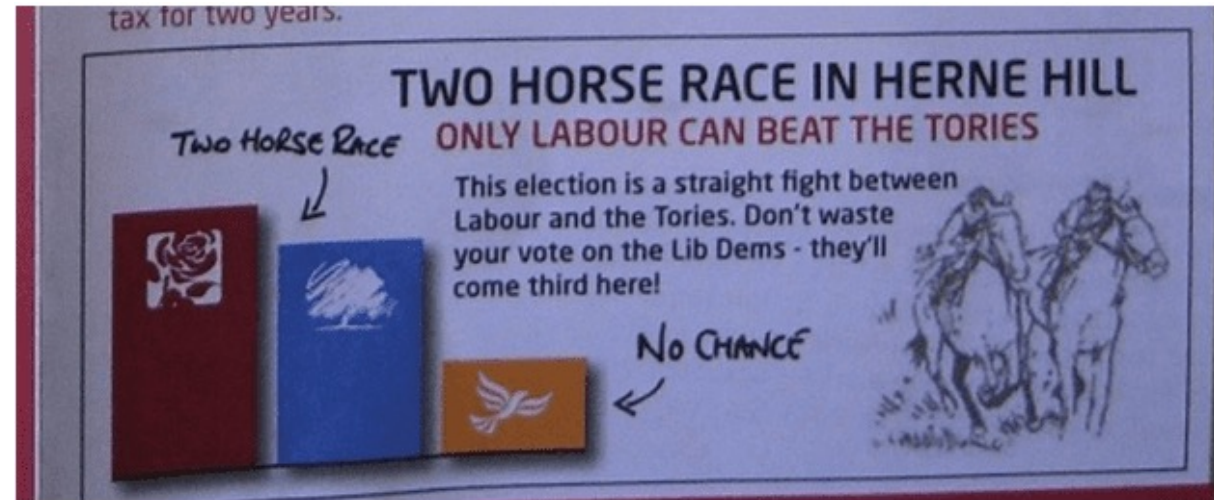
- Omitting baseline
- **Misleading scale**
- Omitting data
- Breaking conventions
- Distorting reality
- Too many stories
- Too many dimensions
- Too many elements



Common sense & cultural context matter!

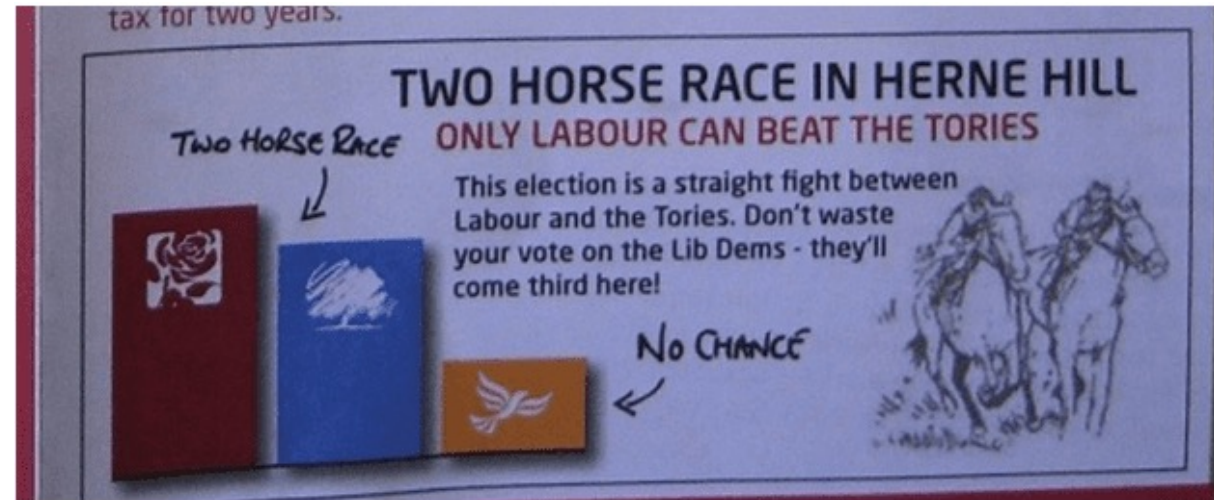
4. Common Pitfalls

- Omitting baseline
- Misleading scale
- **Omitting data**
- Breaking conventions
- Distorting reality
- Too many stories
- Too many dimensions
- Too many elements

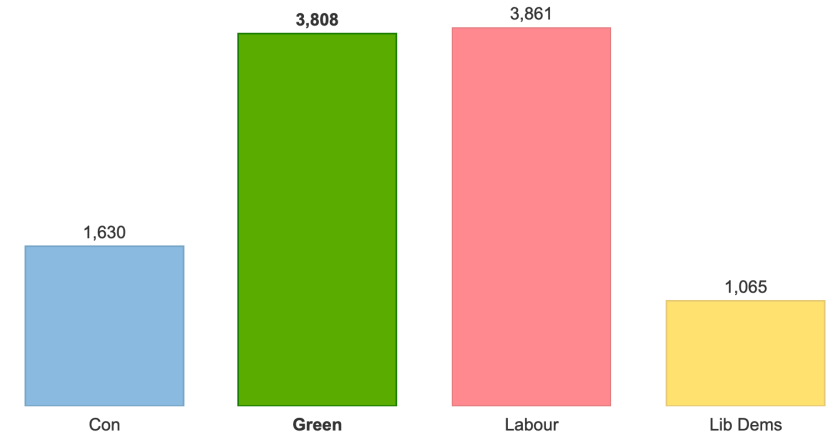


4. Common Pitfalls

- Omitting baseline
- Misleading scale
- **Omitting data**
- Breaking conventions
- Distorting reality
- Too many stories
- Too many dimensions
- Too many elements



Votes by party, Herne Hill ward, Lambeth 2006



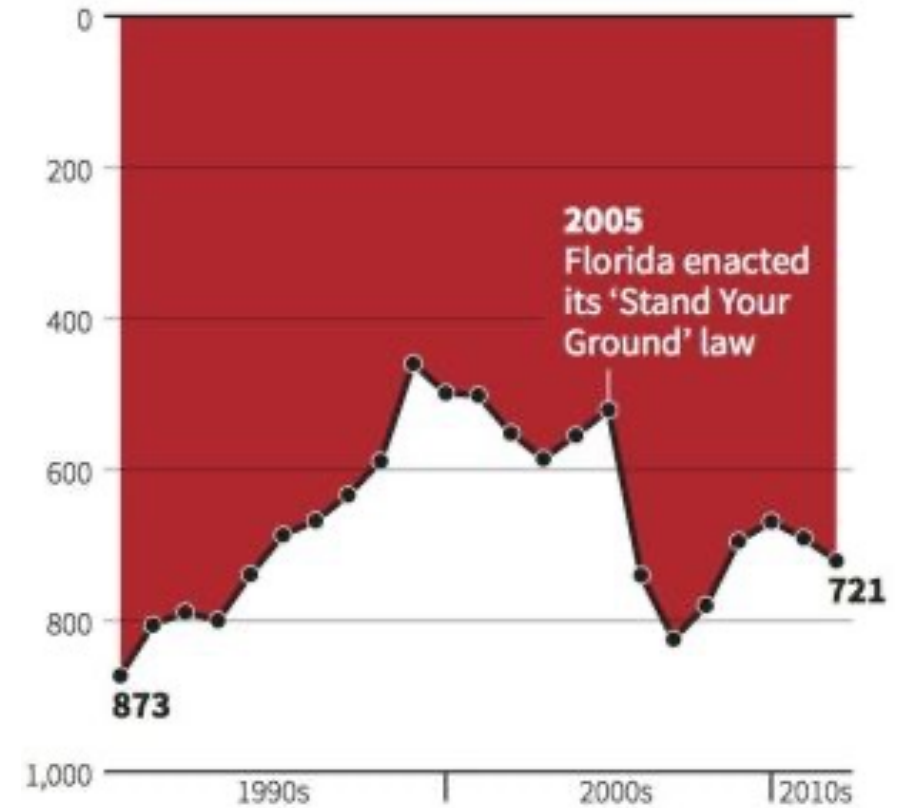
SOURCE: LAMBETH COUNCIL
[GET THE DATA](#) [EMBED](#) [FULLSCREEN](#)

4. Common Pitfalls

- Omitting baseline
- Misleading scale
- Omitting data
- **Breaking conventions**
- Distorting reality
- Too many stories
- Too many dimensions
- Too many elements

Gun deaths in Florida

Number of murders committed using firearms



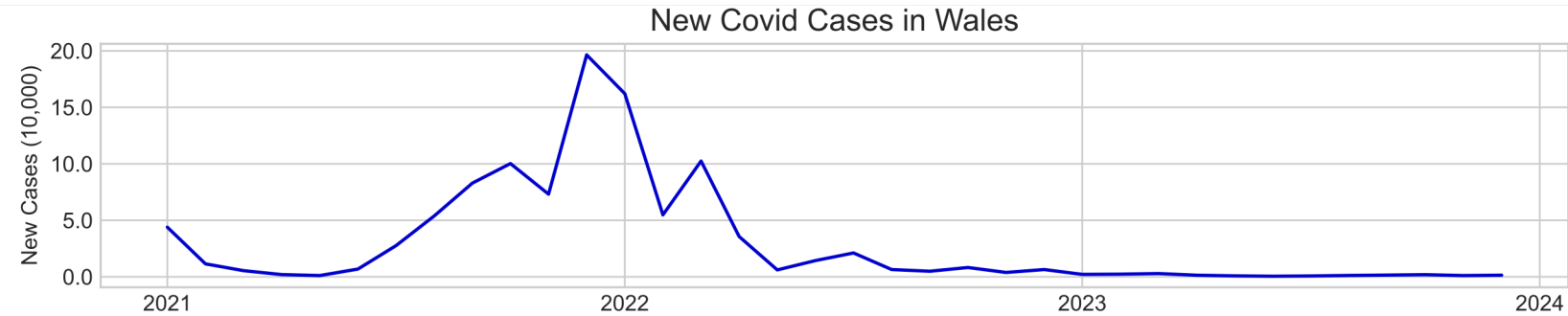
Source: Florida Department of Law Enforcement

C. Chan 16/02/2014

REUTERS

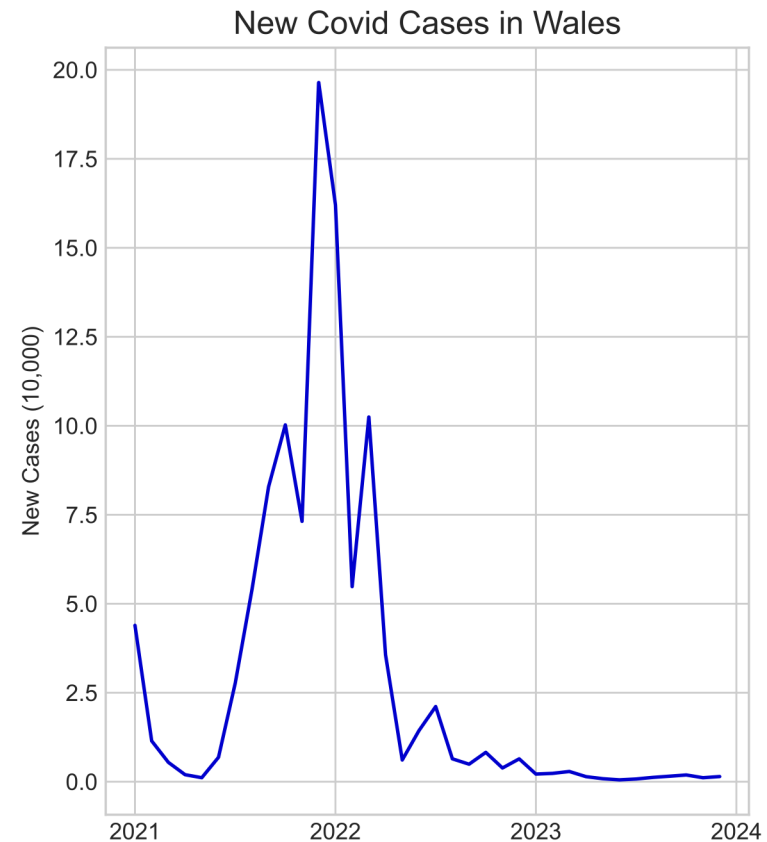
4. Common Pitfalls

- Omitting baseline
- Misleading scale
- Omitting data
- Breaking conventions
- **Distorting reality**
- Too many stories
- Too many dimensions
- Too many elements



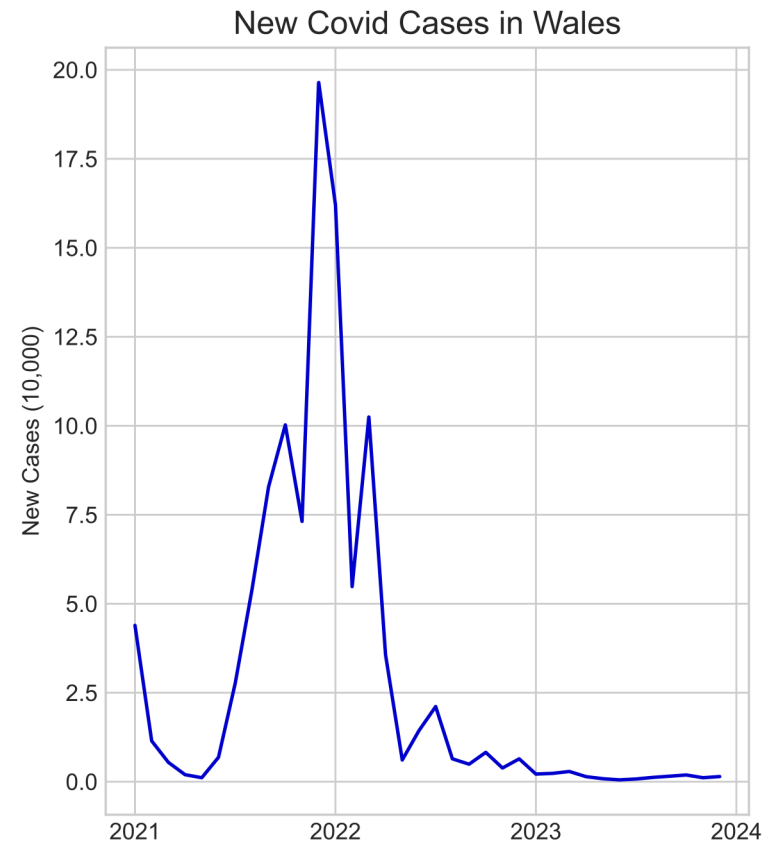
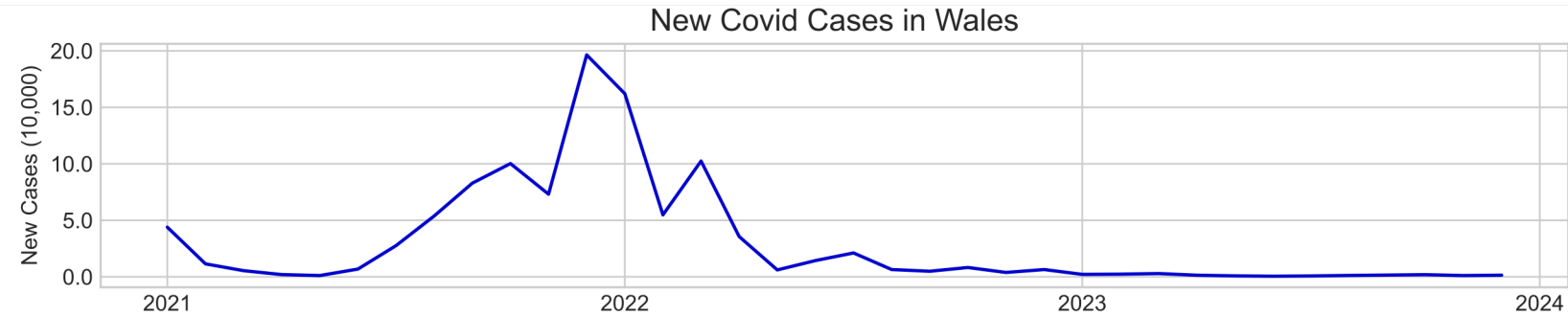
4. Common Pitfalls

- Omitting baseline
- Misleading scale
- Omitting data
- Breaking conventions
- **Distorting reality**
- Too many stories
- Too many dimensions
- Too many elements



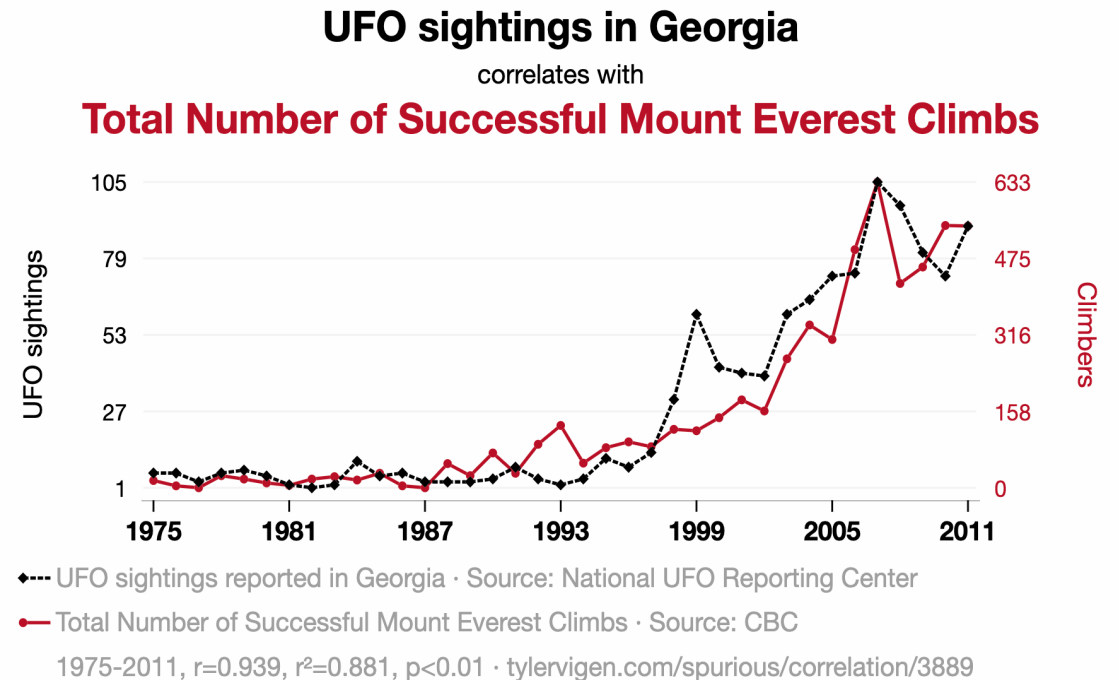
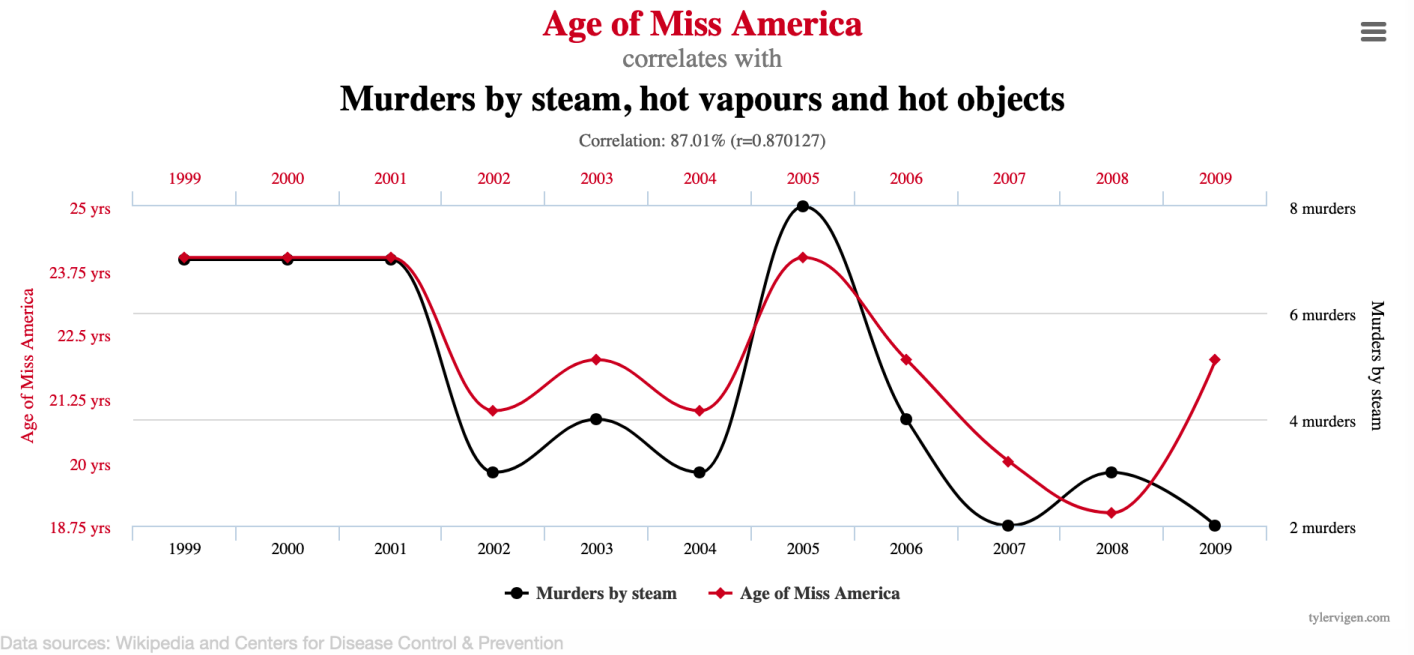
4. Common Pitfalls

- Omitting baseline
- Misleading scale
- Omitting data
- Breaking conventions
- **Distorting reality**
- Too many stories
- Too many dimensions
- Too many elements



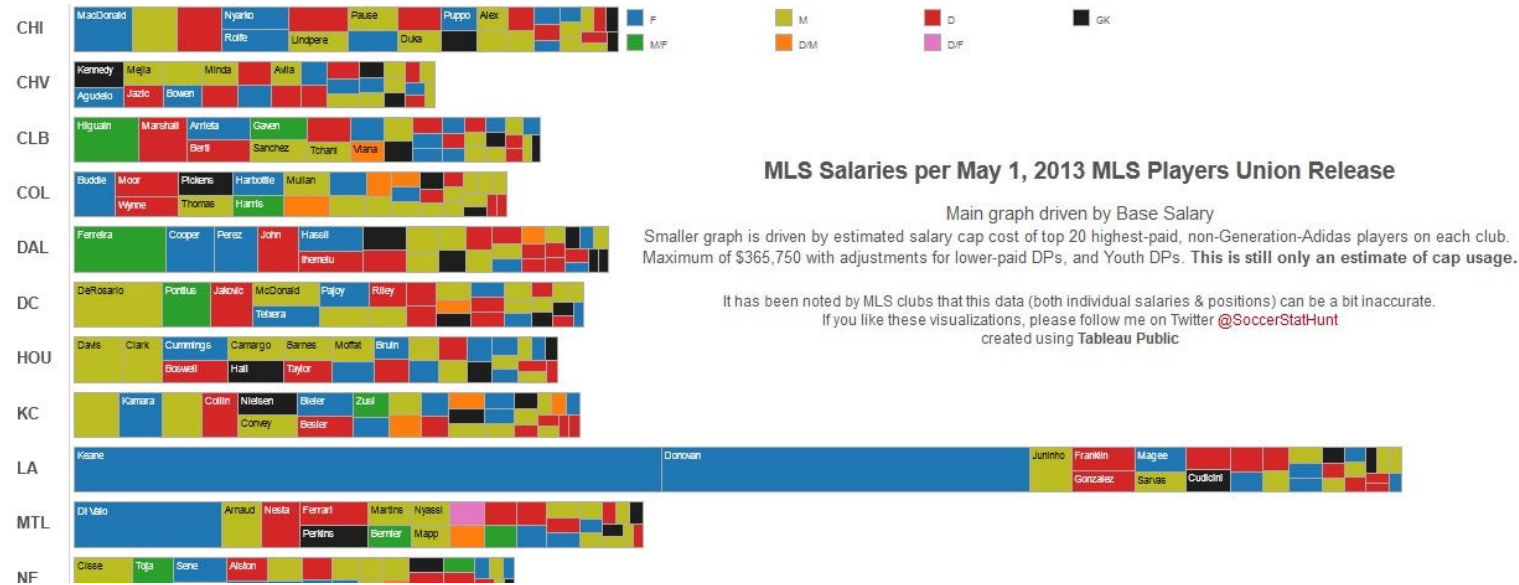
4. Common Pitfalls

- Omitting baseline
- Misleading scale
- Omitting data
- Breaking conventions
- Distorting reality
- **Too many stories**
- Too many dimensions
- Too many elements



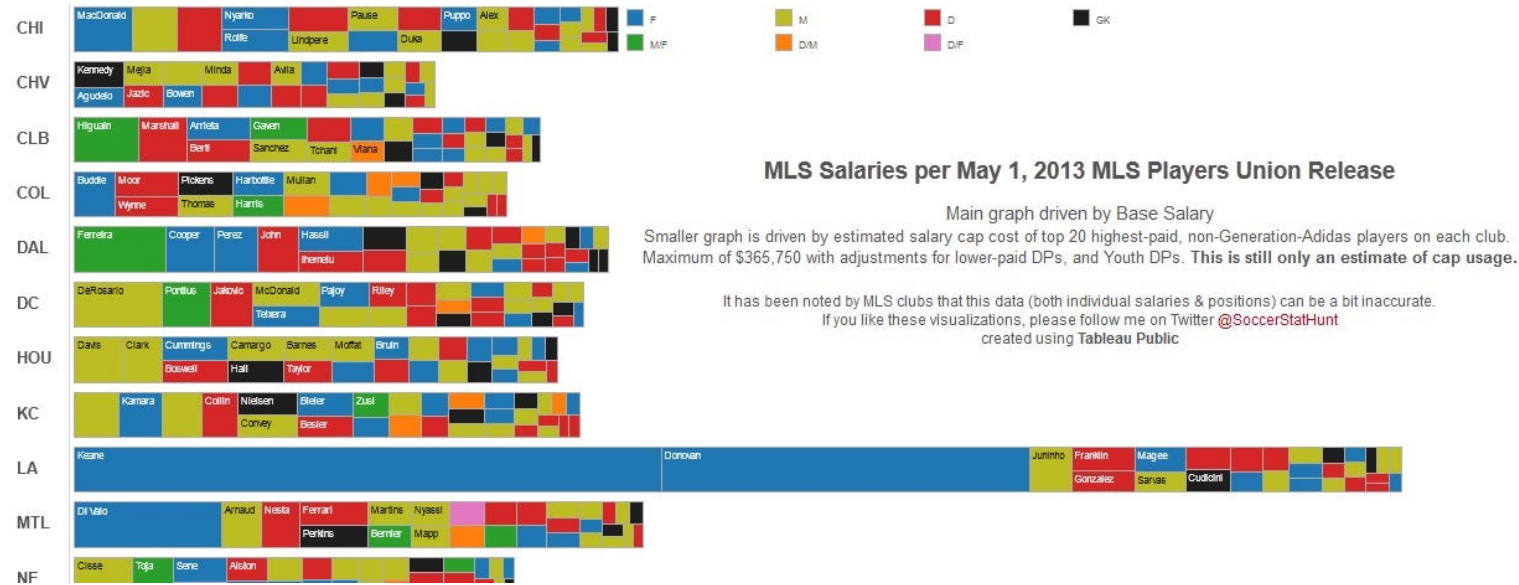
4. Common Pitfalls

- Omitting baseline
- Misleading scale
- Omitting data
- Breaking conventions
- Distorting reality
- Too many stories
- **Too many dimensions**
- Too many elements



4. Common Pitfalls

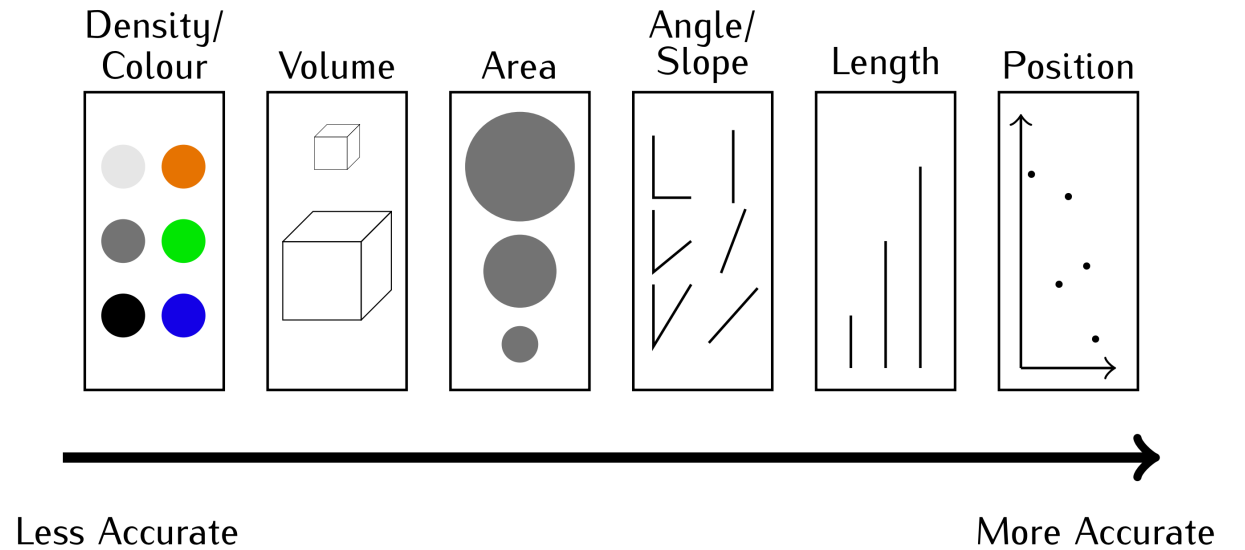
- Omitting baseline
- Misleading scale
- Omitting data
- Breaking conventions
- Distorting reality
- Too many stories
- **Too many dimensions**
- Too many elements



Rule of thumb: stick to 1, 2, or at most 3 variables!

4. Common Pitfalls

- Omitting baseline
- Misleading scale
- Omitting data
- Breaking conventions
- Distorting reality
- Too many stories
- Too many dimensions
- **Too many elements**



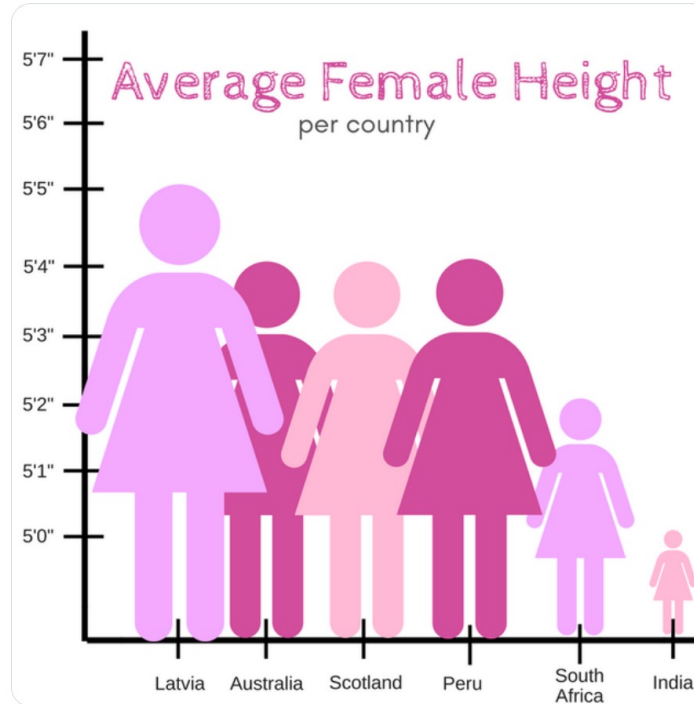
4. Common Pitfalls

- Omitting baseline
- Misleading scale
- Omitting data
- Breaking conventions
- Distorting reality
- Too many stories
- Too many dimensions
- Too many elements



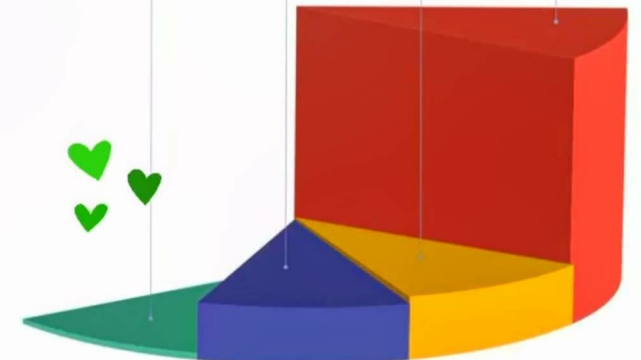
Sabah Ibrahim
@reina_sabah

As an Indian woman, I can confirm that too much of my time is spent hiding behind a rock praying the terrifying gang of international giant ladies and their Latvian general don't find me



11:58 pm · 6 Aug 2020 · Twitter Web App

2021 EUROPEAN YEAR OF RAIL



Greenhouse gas emissions from transport in the EU (2017)

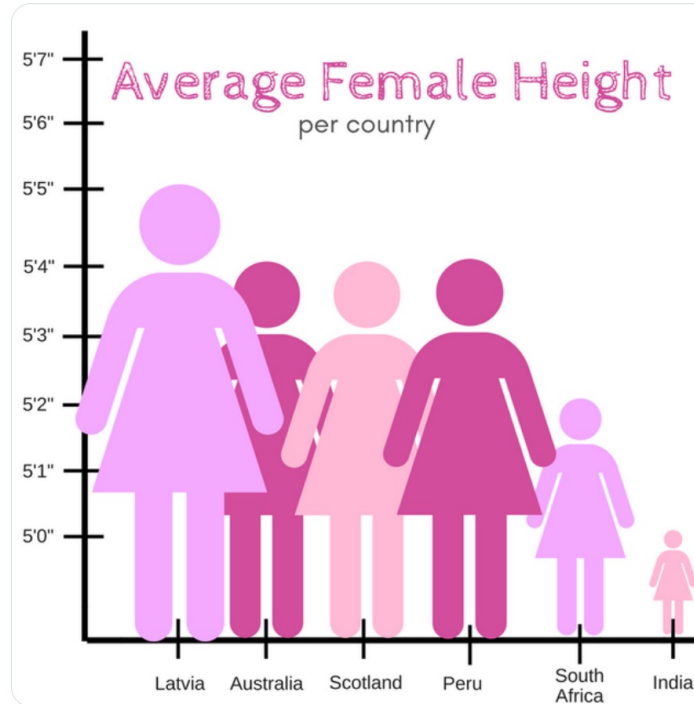
4. Common Pitfalls

- Omitting baseline
- Misleading scale
- Omitting data
- Breaking conventions
- Distorting reality
- Too many stories
- Too many dimensions
- Too many elements



Sabah Ibrahim
@reina_sabah

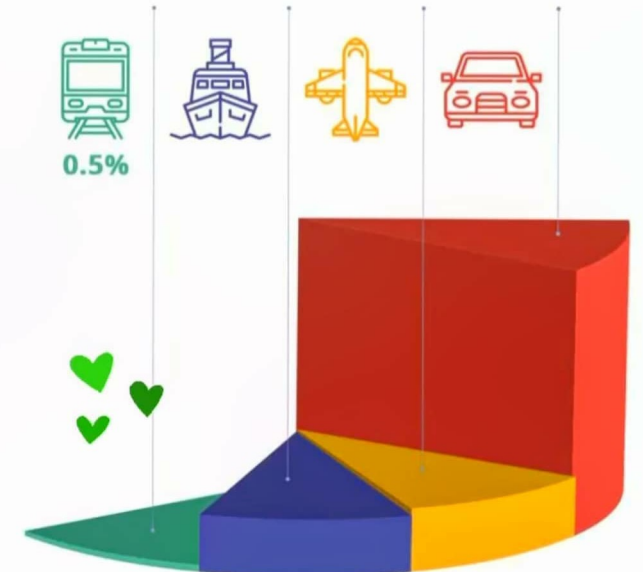
As an Indian woman, I can confirm that too much of my time is spent hiding behind a rock praying the terrifying gang of international giant ladies and their Latvian general don't find me



11:58 pm · 6 Aug 2020 · Twitter Web App

2021 EUROPEAN YEAR OF RAIL

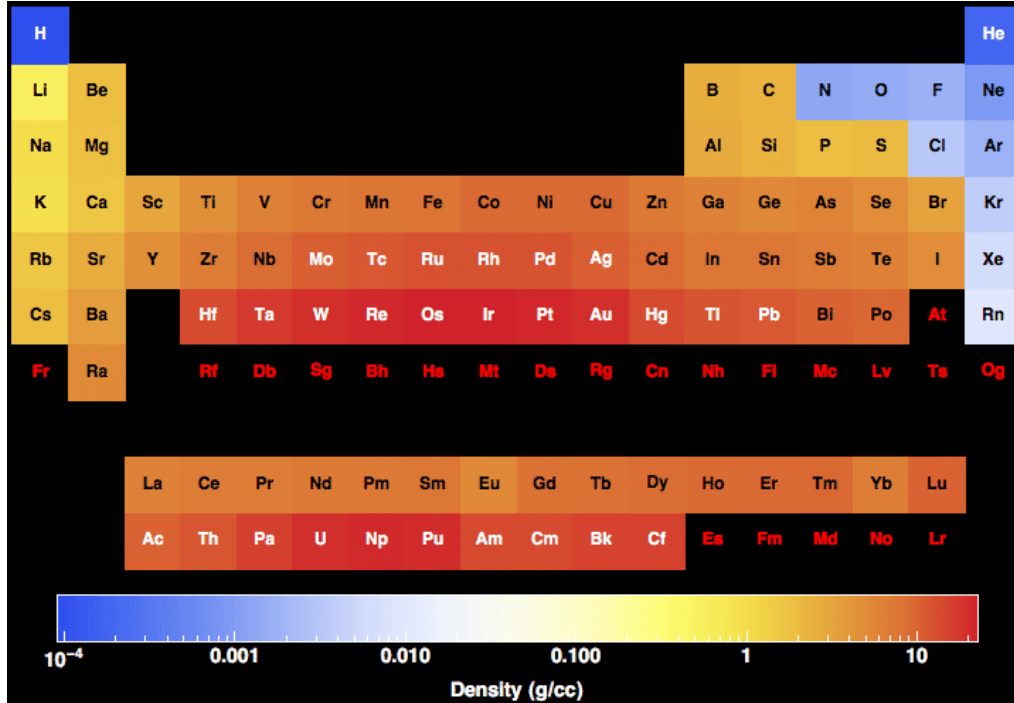
0.5%



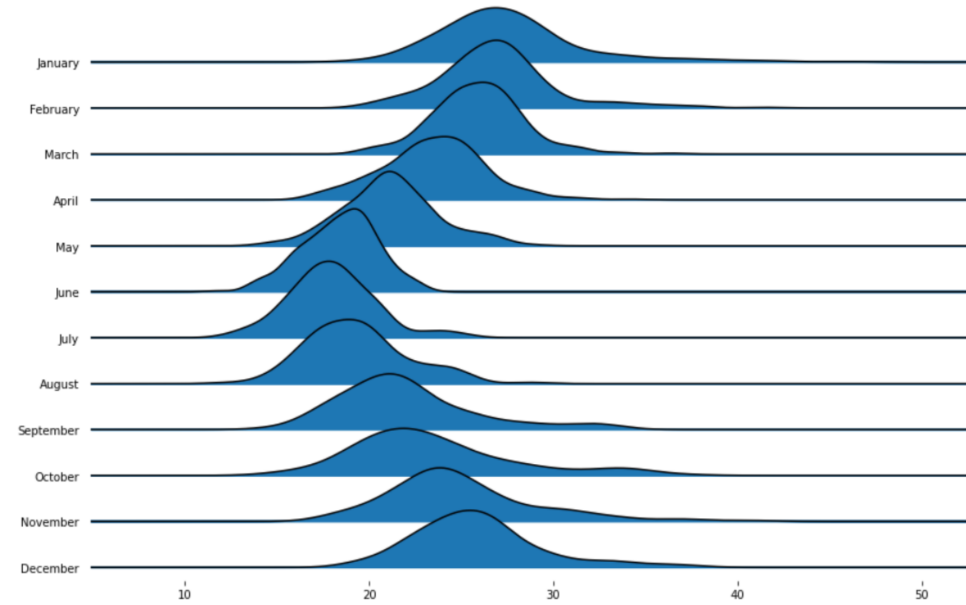
Greenhouse gas emissions from transport in the EU (2017)

Rule of thumb: 1 element per variable!

4. Creativity

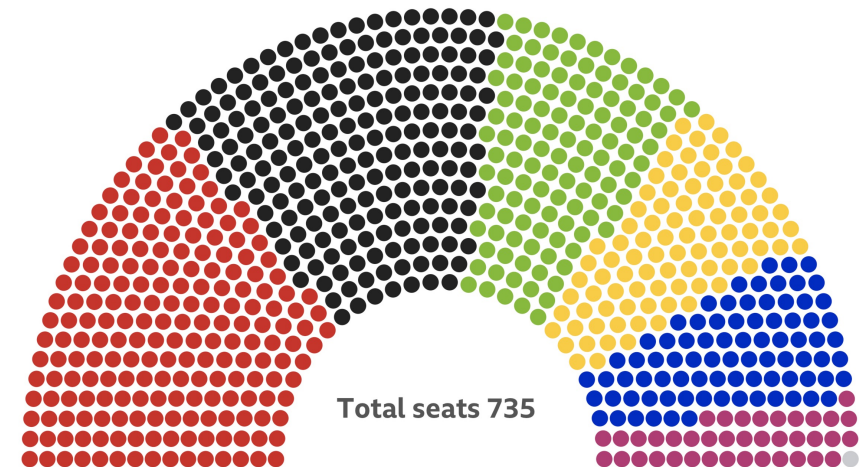


Ridgeline Plot of Max Temperatures in Sydney



How the new German parliament looks

- SPD (206) ● CDU-CSU (196) ● Greens (118) ● FDP (92)
- AfD (83) ● Linke (39) ● SSW (1)



4. Caution! Creativity

Clarity
Understanding
Objectivity



Beauty
Cleverness

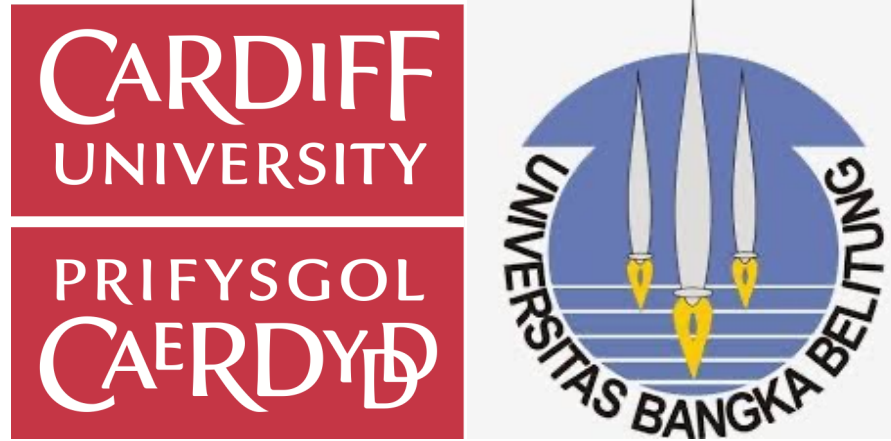
4. Caution! Creativity

Clarity
Understanding
Objectivity



Beauty
Cleverness

We have standard tools and
rules to get these right!



Questions?

Dr Geraint Palmer-Liyu

School of Mathematics - Cardiff University

FEB UBB – Cardiff University: International Visiting Lecture Series